Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/GB04/005435

International filing date: 23 December 2004 (23.12.2004)

Document type: Certified copy of priority document

Document details: Country/Office: GB
Number: 0329722.3
Filing date: 23 December 2003 (23.12.2003)

Date of receipt at the International Bureau: 17 March 2005 (17.03.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

PCT|GB 2004|005435

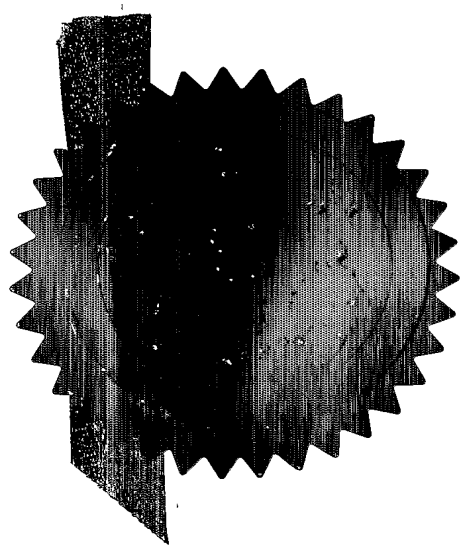
The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed Andrew Gerry

Dated 10 February 2005

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

23DEC03 E861492-1 002866
P01/7700 0.00-0329722.3 ACCOUNT CHA

1/77

THE PATENT OFFICE
SN
23 DEC 2003
RECEIVED BY FAX

The Patent Office

Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference — DELE / P25845GB

2. Patent application number *(The Patent Office will fill in this part)* — 0329722.3

3. Full name, address and postcode of the or of each applicant *(underline all surnames)* — **Delta Biotechnology ~~Ltd~~** LIMITED
Castle Court
59 Castle Boulevard
Nottingham NG7 1FD
United Kingdom

Patents ADP number *(if you know it)* — 56777290 01

If the applicant is a corporate body, give the country/state of its incorporation — United Kingdom

AC 2.4.04

4. Title of the invention — MODIFIED PLASMID AND USE THEREOF

5. Name of your agent *(if you have one)* — **ERIC POTTER CLARKSON**
PARK VIEW HOUSE
58 THE ROPEWALK
NOTTINGHAM
NG1 5DD

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

Patents ADP number *(if you know it)* — **1305010** ✓

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) *any applicant named in part 3 is not an inventor; or*
b) *there is an inventor who is not named as an applicant, or*
c) *any named applicant is a corporate body. See note (d))*

YES

0089820 23-Dec-03 12:49

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	
Description	67
Claims(s)	10
Abstract	1
Drawing(s)	39

SN

10. If you are also filing in any of the following, state how many against each item.

Priority Documents	0
Translations of priority documents	0
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	NO
Request for preliminary examination and search *(Patents Form 9/77)*	YES
Request for substantive examination *(Patents Form 10/77)*	NO
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Eric Potter Clark

Signature: **ERIC POTTER CLARKSON** Date: **23 December 2003**

12. Name and daytime telephone number of person to contact in the United Kingdom: **0115 9552211**

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 01645 500505.*
b) *Write your answers in capital letters using black ink or you may type them.*
c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*
d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*
e) *Once you have filled in the form you must remember to sign and date it.*
f) *For details of the fee and ways to pay please contact the Patent Office.*

MODIFIED PLASMID AND USE THEREOF

FIELD OF THE INVENTION

The present application relates to modified plasmids and uses thereof.

BACKGROUND OF THE INVENTION

Certain closely related species of budding yeast have been shown to contain naturally occurring circular double stranded DNA plasmids. These plasmids, collectively termed 2μm-family plasmids, include pSR1, pSB3 and pSB4 from *Zygosaccharomyces rouxii* (formerly classified as *Zygosaccharomyces bisporus*), plasmids pSB1 and pSB2 from *Zygosaccharomyces bailii*, plasmid pSM1 from *Zygosaccharomyces fermentati*, plasmid pKD1 from *Kluyveromyces drosphilarum*, an un-named plasmid from *Pichia membranaefaciens* (hereinafter referred to as "pPM1") and the 2μm plasmid and variants (such as Scp1, Scp2 and Scp3) from *Saccharomyces cerevisiae* (Volkert, *et al.*, 1989, *Microbiological Reviews*, **53**, 299; Painting, *et al.*, 1984, *J. Applied Bacteriology*, **56**, 331) and other *Saccharomyces* species, such as S. *carlsbergensis*. As a family of plasmids these molecules share a series of common features in that they possess two inverted repeats on opposite sides of the plasmid, have a similar size around 6-kbp (range 4757 to 6615-bp), at least three open reading frames, one of which encodes for a site specific recombinase (such as *FLP* in 2μm) and an autonomously replicating sequence (*ARS*), also known as an origin of replication (*ori*), located close to the end of one of the inverted repeats. (Futcher, 1988, *Yeast*, **4**, 27; Murray *et al.*, 1988, *J. Mol. Biol.* **200**, 601 and Toh-e *et al.*, 1986, *Basic Life Sci.* **40**, 425). Despite their lack of discernible DNA sequence homology, their shared molecular architecture and the conservation of function of the open reading frames have demonstrated a common link between the family members.

The 2μm plasmid (Figure 1) is a 6,318-bp double-stranded DNA plasmid, endogenous in most *Saccharomyces cerevisiae* strains at 60-100 copies per haploid genome. The 2μm plasmid comprises a small unique (US) region and a large unique (UL) region, separated by two 599-bp inverted repeat sequences. Site-specific recombination of the inverted repeat sequences results in inter-conversion between the A-form and B-form of the plasmid *in vivo* (Volkert & Broach, 1986, *Cell*, 46, 541). The two forms of 2μm differ only in the relative orientation of their unique regions.

While DNA sequencing of a cloned 2μm plasmid (also known as Scp1) from *Saccharomyces cerevisiae* gave a size of 6,318-bp (Hartley and Donelson, 1980, *Nature*, 286, 860), other slightly smaller variants of 2μm, Scp2 and Scp3, are known to exist as a result of small deletions of 125-bp and 220-bp, respectively, in a region known as *STB* (Cameron *et al.*, 1977, *Nucl. Acids Res.*, 4, 1429; Kikuchi, 1983, *Cell*, 35, 487 and Livingston & Hahne, 1979, *Proc. Natl. Acad. Sci. USA*, 76, 3727). In one study about 80% of natural *Saccharomyces* strains from around the world contained DNA homologous to 2μm (by Southern blot analysis) (Hollenberg, 1982, *Current Topics in Microbiology and Immunobiology*, 96, 119). Furthermore, variation (genetic polymorphism) occurs within the natural population of 2μm plasmids found in *S. cerevisiae* and *S. carlsbergensis*, with the NCBI sequence (accession number NC_001398) being one example.

The 2μm plasmid has a nuclear localisation and displays a high level of mitotic stability (Mead *et al*, 1986, *Molecular & General Genetics*, 205, 417). The inherent stability of the 2μm plasmid results from a plasmid-encoded copy number amplification and partitioning mechanism, which is easily compromised during the development of chimeric vectors (Futcher & Cox, 1984, *J. Bacteriol.*, 157, 283; Bachmair & Ruis, 1984, *Monatshefte für Chemie*, 115, 1229). A yeast strain,

which contains a 2μm plasmid is known as [cir$^+$], while a yeast strain which does not contain a 2μm plasmid is known as [cir^0].

The US-region contains the *REP2* and *FLP* genes, and the UL-region contains the *REP1* and *D* (also known as *RAF*) genes, the *STB*-locus and the origin of replication (Broach & Hicks, 1980, *Cell*, **21**, 501; Sutton & Broach, 1985, *Mol. Cell. Biol.*, **5**, 2770). The Flp recombinase binds to FRT-sites (Flp Recognition Target) within the inverted repeats to mediate site-specific recombination, which is essential for natural plasmid amplification and control of plasmid copy number *in vivo* (Senecoff *et al*, 1985, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 7270; Jayaram, 1985, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 5875). The copy number of 2μm-family plasmids can be significantly affected by changes in Flp recombinase activity (Sleep *et al*, 2001, *Yeast*, **18**, 403; Rose & Broach, 1990, *Methods Enzymol.*, **185**, 234). The Rep1 and Rep2 proteins mediate plasmid segregation, although their mode of action is unclear (Sengupta *et al*, 2001, *J. Bacteriol.*, **183**, 2306). They also repress transcription of the *FLP* gene (Reynolds *et al*, 1987, *Mol. Cell. Biol.*, **7**, 3566).

The *FLP* and *REP2* genes are transcribed from divergent promoters, with apparently no intervening sequence defined between them. The *FLP* and *REP2* transcripts both terminate at the same sequence motifs within the inverted repeat sequences, at 24-bp and 178-bp respectively after their translation termination codons (Sutton & Broach, 1985, *Mol. Cell. Biol.*, **5**, 2770).

In the case of *FLP*, the C-terminal coding sequence also lies within the inverted repeat sequence. Furthermore, the two inverted repeat sequences are highly conserved over 599-bp, a feature considered advantageous to efficient plasmid replication and amplification *in vivo*, although only the FRT-sites (less than 65-bp) are essential for site-specific recombination *in vitro* (Senecoff *et al*, 1985, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 7270; Jayaram, 1985, *Proc. Natl. Acad. Sci. U.S.A.*, **82**,

5875; Meyer-Leon *et al*, 1984, *Cold Spring Harbor Symposia On Quantitative Biology*, 49, 797). The key catalytic residues of Flp are arginine-308 and tyrosine-343 (which is essential) with strand-cutting facilitated by histidine-309 and histidine 345 (Prasad *et al*, 1987, *Proc. Natl. Acad. Sci. U.S.A.*, 84, 2189; Chen *et al*, 1992, *Cell*, 69, 647; Grainge *et al*, 2001, *J. Mol. Biol.*, 314, 717).

Two functional domains are described in Rep2. Residues 15-58 form a Rep1-binding domain, and residues 59-296 contain a self-association and STB-binding region (Sengupta *et al*, 2001, *J. Bacteriol.*, 183, 2306).

Chimeric or large deletion mutant derivatives of 2μm which lack many of the essential functional regions of the 2μm plasmid but retain functional the *cis* element *ARS* and *STB*, cannot effectively partition between mother and daughter cells at cell division. Such plasmids can do so if these functions are supplied in *trans*, by for instance the provision of a functional 2μm plasmid within the host, a so called [cir^+] host.

Genes of interest have previously been inserted into the UL-region of the 2μm plasmid. For example, see plasmid pSAC3U1 in EP 0 286 424. However, there is likely to be a limit to the amount of DNA that can usefully be inserted into the UL-region of the 2μm plasmid without generating excessive asymmetry between the US and UL-regions. Therefore, the US-region of the 2μm plasmid is particularly attractive for the insertion of additional DNA sequences, as this would tend to equalise the length of DNA fragments either side of the inverted repeats.

This is especially true for expression vectors, such as that shown in Figure 2, in which the plasmid is already crowded by the introduction of a yeast selectable marker and adjacent DNA sequences. For example, the plasmid shown in Figure 2 includes a β-lactamase gene (for ampicillin resistance), a *LEU2* selectable marker and an oligonucleotide linker, the latter two of which are inserted into a

unique *Sna*BI-site within the UL-region of the 2μm-family disintegration vector, pSAC3 (see EP 0 286 424). The *E. coli* DNA between the *Xba*I-sites that contains the ampicillin resistance gene is lost from the plasmid shown in Figure 2 after transformation into yeast. This is described in Chinery & Hinchliffe, 1989, *Curr. Genet.*, 16, 21 and EP 0 286 424, where these types of vectors are designated "disintegration vectors". In the crowded state shown in Figure 2, it is not readily apparent where further polynucleotide insertions can be made. A *Not*I-site within the linker has been used for the insertion of additional DNA fragments, but this contributes to further asymmetry between the UL and US regions (Sleep *et al*, 1991, *Biotechnology (N Y)*, 9, 183).

We had previously attempted to insert additional DNA into the US-region of the 2μm plasmid and maintain its high inherent plasmid stability. In the 2μm-family disintegration plasmid pSAC300, a 1.1-kb DNA fragment containing the *URA3* gene was inserted into *Eag*I-site between *REP2* and *FLP* in US-region in such a way that transcription from the *URA3* gene was in same direction as *REP2* transcription (see EP 0 286 424). When S150-2B [cir^0] was transformed to uracil prototrophy by pSAC300, it was shown to be considerably less stable (50% plasmid loss in under 30 generations) than comparable vectors with *URA3* inserted into the UL-region of 2μm (0-10% plasmid loss in under 30 generations) (Chinery & Hinchliffe, 1989, *Curr. Genet.*, 16, 21; EP 0 286 424). Thus, insertion at the *Eag*I site may have interfered with *FLP* expression and it was concluded that the insertion position could have a profound effect upon the stability of the resultant plasmid a conclusion confirmed by Bijvoet *et al.*, 1991, *Yeast*, 7, 347.

It is desirable to insert further polynucleotide sequences into 2μm-family plasmids. For example, the insertion of polynucleotide sequences that encode host derived proteins, recombinant proteins, or non-coding antisense or RNA interference (RNAi) transcripts may be desirable. Moreover, it is desirable to introduce multiple further polynucleotide sequences into 2μm-family plasmids,

thereby to provide a plasmid which encodes, for example, multiple separately encoded multi-subunit proteins, different members of the same metabolic pathway, additional selective markers or a recombinant protein (single or multi-subunit) and a chaperone to aid the expression of the recombinant protein.

However, the 6,318-bp 2μm plasmid, and other 2μm-family plasmids, are crowded with functional genetic elements (Sutton & Broach, 1985, *Mol. Cell. Biol.*, **5**, 2770; Broach *et al*, 1979, *Cell*, **16**, 827), with no obvious positions existing for the insertion of additional DNA sequences without a concomitant loss in plasmid stability. In fact, except for the region between the origin of replication and the *D* gene locus, the entire 2μm plasmid genome is transcribed into at least one poly(A)$^+$ species and often more (Sutton & Broach, 1985, *Mol. Cell. Biol.*, **5**, 2770). Consequently, most insertions might be expected to have a detrimental impact on plasmid function *in vivo*.

Indeed, persons skilled in the art have given up on inserting heterologous polynucleotide sequences into 2μm-family plasmids.

Robinson *et al*, 1994, *Bio/Technology*, **12**, 381-384 reported that a recombinant additional PDI gene copy in *Saccharomyces cerevisiae* could be used to increase the recombinant expression of human platelet derived growth factor (PDGF) B homodimer by ten-fold and *Schizosacharomyces pombe* acid phosphatase by four-fold. Robinson obtained the observed increases in expression of PDGF and *S. pombe* acid phosphatase using an additional chromosomally integrated PDI gene copy. Robinson reported that attempts to use the multi-copy 2μm expression vector to increase PDI protein levels had had a detrimental effect on heterologous protein secretion.

Shusta *et al*, 1998, *Nature Biotechnology*, **16**, 773-777 described the recombinant expression of single-chain antibody fragments (scFv) in *Saccharomyces*

cerevisiae. Shusta reported that in yeast systems, the choice between integration of a transgene into the host chromosome versus the use of episomal expression vectors can greatly affect secretion and, with reference to Parekh & Wittrup, 1997, *Biotechnol. Prog.*, 13, 117-122, that stable integration of the scFv gene into the host chromosome using a δ integration vector was superior to the use of a 2μm-based expression plasmid. Parekh & Wittrup, *op. cit.*, had previously taught that the expression of bovine pancreatic trypsin inhibitor (BPTI) was increased by an order of magnitude using a δ integration vector rather than a 2μm-based expression plasmid. The 2μm-based expression plasmid was said to be counter-productive for the production of heterologous secreted protein.

Bao *et al*, 2000, *Yeast*, 16, 329-341, reported that the *KlPDI1* gene had been introduced into *K. lactis* an a multi-copy plasmid, pKan707, and that the presence of the plasmid caused the strain to grow poorly. In the light of the earlier findings in Bao *et al*, 2000, Bao & Fukuhara, 2001, *Gene*, 272, 103-110, chose to introduce a single duplication of *KlPDI1* on the host chromosome.

Accordingly, the art teaches the skilled person to integrate transgenes into the yeast chromosome, rather then into a multicopy vector. There is, therefore, a need for alternative ways of transforming yeast.

DESCRIPTION OF THE INVENTION

The present invention relates to recombinantly modified versions of 2μm-family plasmids.

A 2μm-family plasmid is a circular, double stranded, DNA plasmid. It is typically small, such as between 3,000 to 10,000 bp, preferably between 4,500 to 7000 bp, excluding recombinantly inserted sequences. Preferred 2μm-family plasmids for use in the present invention comprise sequences derived from one or more of

plasmids pSR1, pSB3, or pSB4 as obtained from *Zygosaccharomyces rouxii*, pSB1 or pSB2 both as obtained from *Zygosaccharomyces bailli*, pSM1 as obtained from *Zygosaccharomyces fermentati*, pKD1 as obtained from *Kluyveromyces drosophilarum*, pPM1 as obtained from *Pichia membranaefaciens* and the 2μm plasmid and variants (such as Scp1, Scp2 and Scp3) as obtained from *Saccharomyces cerevisiae*, for example as described in Volkert *et al*, 1989, *Microbiological Reviews*, **53**(3), 299-317, Murray *et al*, 1988, *Mol. Biol.*, **200**, 601-607 and Painting, *et al.*, 1984, *J. Applied Bacteriology*, **56**, 331.

A 2μm-family plasmid is capable of stable multicopy maintenance within a yeast population, although not necessarily all 2μm-family plasmids will be capable of stable multicopy maintenance within all types of yeast population. For example, the 2μm plasmid is capable of stable multicopy maintenance, *inter alia*, within *Saccharomyces cerevisiae* and *Saccharomyces carlsbergensis*.

By "multicopy maintenance" we mean that the plasmid is present in multiple copies within each yeast cell. A yeast cell comprising 2μm-family plasmid is designated [cir^+], whereas a yeast cell that does not comprise 2μm-family plasmid is designated [cir^0]. A [cir^+] yeast cell typically comprises 10-100 copies of 2μm-family plasmid per haploid genome, such as 20-90, more typically 30-80, preferably 40-70, more preferably 50-60 copies per haploid genome. Moreover, the plasmid copy number can be affected by the genetic background of the host which can increase the plasmid copy number of 2μm-like plasmid to above 100 per haploid genome (Gerbaud and Guerineau, 1980, *Curr. Genetics*, **1**, 219, Holm, 1982, *Cell*, **29**, 585, Sleep *et al.*, 2001, *Yeast*, **18**, 403 and WO99/00504). Multicopy stability is defined below. .

A 2μm-family plasmid typically comprises at least three open reading frames ("ORFs") that each encodes a protein that functions in the stable maintenance of

the 2μm-family plasmid as a multicopy plasmid. The proteins encoded by the three ORFs can be designated FLP, REP1 and REP2. Where a 2μm-family plasmid comprises not all three of the ORFs encoding FLP, REP1 and REP2 then ORFs encoding the missing protein(s) should be supplied in *trans*, either on another plasmid or by chromosomal integration.

A "FLP" protein is a protein capable of catalysing the site-specific recombination between inverted repeat sequences recognised by FLP. The inverted repeat sequences are termed FLP recombination target (FRT) sites and each is typically present as part of a larger inverted repeat (see below). Preferred FLP proteins comprise the sequence of the FLP proteins encoded by one of plasmids pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid, for example as described in Volkert *et al, op. cit.*, Murray *et al, op. cit* and Painting *et al, op. cit.* Variants and fragments of these FLP proteins are also included in the present invention. "Fragments" and "variants" are those which retain the ability of the native protein to catalyse the site-specific recombination between the same FRT sequences. Such variants and fragments will usually have at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or more, homology with an FLP protein encoded by one of plasmids pSR1, pSB1, pSB2, pSB3, pSM1, pKD1 and the 2μm plasmid. Different FLP proteins can have different FRT sequence specificities. A typical FRT site may comprise a core nucleotide sequence flanked by inverted repeat sequences. In the 2μm plasmid, the FRT core sequence is 8 nucleotides in length and the flanking inverted repeat sequences are 13 nucleotides in length (Volkert *et al, op. cit.*). However the FRT site recognised by any given FLP protein may be different to the 2μm plasmid FRT site.

REP1 and REP2 are proteins involved in the partitioning of plasmid copies during cell division, and may also have a role in the regulation of FLP expression. Considerable sequence divergence has been observed between REP1 proteins from different 2μm-family plasmids, whereas no sequence alignment is currently

possible between REP2 proteins derived from different 2μm-family plasmids. Preferred REP1 and REP2 proteins comprise the sequence of the REP1 and REP2 proteins encoded by one of plasmids pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid, for example as described in Volkert *et al*, *op. cit.*, Murray *et al*, *op. cit.* and Painting *et al*, *op. cit.* Variants and fragments of these REP1 and REP2 proteins are also included in the present invention. "Fragments" and "variants" of REP1 and REP2 are those which, when encoded by the plasmid in place of the native ORF, do not disrupt the stable multicopy maintenance of the plasmid within a suitable yeast population. Such variants and fragments of REP1 and REP2 will usually have at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or more, homology with a REP1 and REP2 protein, respectively, as encoded by one of plasmids pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid.

The REP1 and REP2 proteins encoded by the ORFs on the plasmid must be compatible. REP1 and REP2 are compatible if they contribute, in combination with the other functional elements of the plasmid, towards the stable multicopy maintenance of the plasmid which encodes them. Whether or not a REP1 and REP2 ORF contributes towards the stable multicopy maintenance of the plasmid which encodes them can be determined by preparing mutants of the plasmid in which each of the REP1 and REP2 ORFs are specifically disrupted. If the disruption of an ORF impairs the stable multicopy maintenance of the plasmid then the ORF can be concluded to contribute towards the stable multicopy maintenance of the plasmid in the non-mutated version. It is preferred that the REP1 and REP2 proteins have the sequences of REP1 and REP2 proteins encoded by the same naturally occurring 2μm-family plasmid, such as pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid, or variant or fragments thereof.

A 2μm-family plasmid comprises two inverted repeat sequences. The inverted repeats may be any size, so long as they each contain an FRT site (see above). The inverted repeats are typically highly homologous. They may share greater than 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% or more sequence identity. In a preferred embodiment they are identical. Typically the inverted repeats are each between 200 to 1000 bp in length. Preferred inverted repeat sequences may each have a length of from 200 to 300 bp, 300 to 400 bp, 400 to 500 bp, 500 to 600 bp, 600 to 700 bp, 700 to 800 bp, 800 to 900 bp, or 900 to 1000 bp. Particularly preferred inverted repeats are those of the plasmids pSR1 (959 bp), pSB1 (675 bp), pSB2 (477 bp), pSB3 (391 bp), pSM1 (352 bp), pKD1 (346 bp), the 2μm plasmid (599 bp), pSB4 and pPM1.

The sequences of the inverted repeats may be varied. However, the sequences of the FRT site in each inverted repeat should be compatible with the specificity of the FLP protein encoded by the plasmid, thereby to enable the encoded FLP protein to act to catalyse the site-specific recombination between the inverted repeat sequences of the plasmid. Recombination between inverted repeat sequences (and thus the ability of the FLP protein to recognise the FRT sites with the plasmid) can be determined by methods known in the art. For example, a plasmid in a yeast cell under conditions that favour FLP expression can be assayed for changes in the restriction profile of the plasmid which would result from a change in the orientation of a region of the plasmid relative to another region of the plasmid. The detection of changes in restriction profile indicate that the FLP protein is able to recognise the FRT sites in the plasmid and therefore that the FRT site in each inverted repeat are compatible with the specificity of the FLP protein encoded by the plasmid.

In a particularly preferred embodiment, the sequences of inverted repeats, including the FRT sites, are derived from the same 2μm-family plasmid as the

ORF encoding the FLP protein, such as pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 or the 2μm plasmid.

The inverted repeats are typically positioned within the 2μm-family plasmid such that the two regions defined between the inverted repeats (e.g. such as defined as UL and US in the 2μm plasmid) are of approximately similar size, excluding exogenously introduced sequences such as transgenes. For example, one of the two regions may have a length equivalent to at least 40%, 50%, 60%, 70%, 80%, 90%, 95% or more, up to 100%, of the length of the other region.

A 2μm-family plasmid comprises the ORF that encodes FLP and one inverted repeat (arbitrarily termed "IR1" to distinguish it from the other inverted repeat mentioned in the next paragraph) juxtaposed in such a manner that IR1 occurs at the distal end of the FLP ORF, without any intervening coding sequence, for example as seen in the 2μm plasmid. By "distal end" in this context we mean the end of the FLP ORF opposite to the end from which the promoter initiates its transcription. In a preferred embodiment, the distal end of the FLP ORF overlaps with IR1.

A 2μm-family plasmid comprises the ORF that encodes REP2 and the other inverted repeat (arbitrarily termed "IR2" to distinguish it from IR1 mentioned in the previous paragraph) juxtaposed in such a manner that IR2 occurs at the distal end of the REP2 ORF, without any intervening coding sequence, for example as seen in the 2μm plasmid. By "distal end" in this context we mean the end of the REP2 ORF opposite to the end from which the promoter initiates its transcription.

In one embodiment, the ORFs encoding REP2 and FLP may be present on the same region of the two regions defined between the inverted repeats of the 2μm-family plasmid, which region may be the bigger or smaller of the regions (if there is any inequality in size between the two regions).

In one embodiment, the ORFs encoding REP2 and FLP may be transcribed from divergent promoters.

Typically, the regions defined between the inverted repeats (e.g. such as defined as UL and US in the 2μm plasmid) of a 2μm-family plasmid may comprise not more than two endogenous genes that encode a protein that functions in the stable maintenance of the 2μm-family plasmid as a multicopy plasmid. Thus in a preferred embodiment, one region of the plasmid defined between the inverted repeats may comprise not more than the ORFs encoding FLP and REP2; FLP and REP1; or REP1 and REP2, as endogenous coding sequence.

A 2μm-family plasmid comprises an origin of replication (also known as an autonomously replicating sequence - "ARS"), which is typically bidirectional. Any appropriate ARS sequence can be present. Consensus sequences typical of yeast chromosomal origins of replication may be appropriate (Broach *et al*, 1982, *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 1165-1174; Williamson, *Yeast*, 1985, **1**, 1-14). Preferred ARSs include those isolated from pSR1, pSB1, pSB2, pSB3, pSB4, pSM1, pKD1, pPM1 and the 2μm plasmid.

Thus, a 2μm-family plasmid typically comprises at least ORFs encoding FLP and REP2, two inverted repeat sequences each inverted repeat comprising an FRT site compatible with FLP protein, and an ARS sequence. Preferably the plasmid also comprises an ORF encoding REP1, although it may be supplied in *trans*, as discussed above. Preferably the FRT sites are derived from the same 2μm-family plasmid as the sequence of the encoded FLP protein. Preferably the sequences of the encoded REP1 and REP2 proteins are derived from the same 2μm-family plasmid as each other. More preferably, the FRT sites are derived from the same 2μm-family plasmid as the sequence of the encoded FLP, REP1 and REP2 proteins. Even more preferably, the sequences of the ORFs encoding FLP, REP1

and REP2, and the sequence of the inverted repeats (including the FRT sites) are derived from the same 2μm-family plasmid. Yet more preferably, the ARS site is obtained from the same 2μm-family plasmid as one or more of the ORFs of FLP, REP1 and REP2, and the sequence of the inverted repeats (including the FRT sites). Preferred plasmids include plasmids pSR1, pSB3 and pSB4 as obtained from *Zygosaccharomyces rouxii*, pSB1 or pSB2 both as obtained from *Zygosaccharomyces bailli*, pSM1 as obtained from *Zygosaccharomyces fermentati*, pKD1 as obtained from *Kluyveromyces drosophilarum*, pPM1 as obtained from *Pichia membranaefaciens*, and the 2μm plasmid as obtained from *Saccharomyces cerevisiae*, for example as described in Volkert *et al*, 1989, *op. cit.*, Murray *et al*, *op. cit.* and Painting *et al*, *op. cit.*

Optionally, a 2μm-family plasmid may comprise a region equivalent to the *STB* region (also known as REP3) of the 2μm plasmid, as defined in Volkert *et al*, *op. cit.* The *STB* region in a 2μm-family plasmid of the invention may comprise two or more tandem repeat sequences, such as three, four, five or more. Alternatively, no tandem repeat sequences may be present. The tandem repeats may be any size, such as 10, 20, 30, 40, 50, 60 70, 80, 90, 100 bp or more in length. The tandem repeats in the *STB* region of the 2μm plasmid are 62 bp in length. It is not essential for the sequences of the tandem repeats to be identical. Slight sequence variation can be tolerated. It may be preferable to select an *STB* region from the same plasmid as either or both of the REP1 and REP2 ORFs. The *STB* region is thought to be a *cis*-acting element and preferably is not transcribed.

Optionally, a 2μm-family plasmid may comprise an additional ORF that encodes a protein that functions in the stable maintenance of the 2μm-family plasmid as a multicopy plasmid. The additional protein can be designated RAF or D. ORFs encoding the RAF or D gene can be seen on, for example, the 2μm plasmid and pSM1. Thus a RAF or D ORF can comprise a sequence suitable to encode the

protein product of the RAF or D gene ORFs encoded by the 2μm plasmid or pSM1, or variants and fragments thereof. Thus variants and fragments of the protein products of the RAF or D genes of the 2μm plasmid or pSM1 are also included in the present invention. "Fragments" and "variants" of the protein products of the RAF or D genes of the 2μm plasmid or pSM1 are those which, when encoded by the 2μm plasmid or pSM1 in place of the native ORF, do not disrupt the stable multicopy maintenance of the plasmid within a suitable yeast population. Such variants and fragments will usually have at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or more, homology with the protein product of the RAF or D gene ORFs encoded by the 2μm plasmid or pSM1.

The present invention provides a 2μm-family plasmid comprising a polynucleotide sequence insertion, deletion and/or substitution between the first base after the last functional codon of at least one of either a *REP2* gene or an *FLP* gene and the last base before the FRT site in an inverted repeat adjacent to said gene.

A polynucleotide sequence insertion is any additional polynucleotide sequence inserted into the plasmid. Preferred polynucleotide sequence insertions are described below. A deletion is removal of one or more base pairs, such as the removal of up to 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more base pairs, which may be as a single contiguous sequence or from spaced apart regions within a DNA sequence. A substitution is the replacement of one or more base pairs, such as the replacement of up to 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more base pairs, which may be as a single contiguous sequence or from spaced apart regions within a DNA sequence. It is possible for a region to be modified by any two of insertion, deletion or substitution, or even all three.

The last functional codon of either a *REP2* gene or a *FLP* gene is the codon in the open reading frame of the gene that is furthest downstream from the promoter of the gene whose replacement by a stop codon will lead to an unacceptable loss of multicopy stability of the plasmid, when determined by a test such as defined in Chinery & Hinchliffe (1989, *Curr. Genet.*, 16, 21-25). For yeast that do not grow in the non-selective media (YPD, also designated YEPD) defined in Chinery & Hinchliffe (1989, *Curr. Genet.*, 16, 21-25) other appropriate non-selective media might be used. Plasmid stability may be defined as the percentage cells remaining prototrophic for the selectable marker after a defined number of generations. The number of generations will preferably be sufficient to show a difference between a control plasmid, such as pSAC35 or pSAC310, or to shown comparable stability to such a control plasmid. The number of generations may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more. Higher numbers are preferred. The acceptable plasmid stability might be 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9% or substantially 100%. Higher percentages are preferred. The skilled person will appreciate that, even though a plasmid may have a stability less than 100% when grown on non-selective media, that plasmid can still be of use when cultured in selective media. For example plasmid pDB2711 as described in the examples is only 10% stable when the stability is determined accordingly to Example 1, but provides a 15-fold increase in recombinant transferrin productivity in shake flask culture under selective growth conditions.

Thus, disruption of the *REP2* or *FLP* genes at any point downstream of the last functional codon in either gene, by insertion of a polynucleotide sequence insertion, deletion or substitution will not lead to an unacceptable loss of multicopy stability of the plasmid. We have surprisingly found that the *REP2* gene of the 2μm plasmid can be disrupted after codon 59 and that the *FLP* gene of the 2μm plasmid can be disrupted after codon 344, each without leading to an

unacceptable loss of multicopy stability of the plasmid. The last functional codon in equivalent genes in other 2μm-family plasmids can be determined routinely by modifying the relevant genes and determining stability as described above.

The *REP2* and *FLP* genes in a 2μm plasmid of the invention each have an inverted repeat adjacent to them. The inverted repeat can be identified because it matches the sequence of another inverted repeat within the same plasmid. By "adjacent" is meant that the *FLP* or *REP2* gene and its inverted repeat are juxtaposed in such a manner that the inverted repeat occurs at the distal end of the gene, without any intervening coding sequence, for example as seen in the 2μm plasmid. By "distal end" in this context we mean the end of the gene opposite to the end from which the promoter initiates its transcription. In a preferred embodiment, the distal end of the gene overlaps with the inverted repeat.

In a first preferred aspect of the invention, the polynucleotide sequence insertion, deletion and/or substitution occurs between the first base after the last functional codon of the *REP2* gene and the last base before the FRT site in an inverted repeat adjacent to said gene, preferably between the first base of the inverted repeat and the last base before the FRT site, even more preferably at a position after the translation termination codon of the *REP2* gene and before the last base before the FRT site.

In a second preferred aspect of the invention, the polynucleotide sequence insertion, deletion and/or substitution occurs between the first base after the last functional codon of the *FLP* gene and the last base before the FRT site in an inverted repeat adjacent to said gene, preferably between the first base of the inverted repeat and the last base before the FRT site, more preferably between the first base after the end of the FLP coding sequence and the last base before the FRT site, such as at the first base after the end of the FLP coding sequence.

In one embodiment, other than the polynucleotide sequence insertion, deletion and/or substitution, the *FLP* gene and/or the *REP2* gene has the sequence of a *FLP* gene and/or a *REP2* gene, respectively, derived from a naturally occurring 2μm-family plasmid.

The term "derived from" includes sequences having an identical sequence to the sequence from which they are derived. However, variants and fragments thereof, as defined above, are also included. For example, an *FLP* gene having a sequence derived from the *FLP* gene of the 2μm plasmid may have a modified promoter or other regulatory sequence compared to that of the naturally occurring gene. Alternatively, an *FLP* gene having a sequence derived from the *FLP* gene of the 2μm plasmid may have a modified nucleotide sequence in the open reading frame which may encode the same protein as the naturally occurring gene, or may encode a modified FLP protein. The same considerations apply to *REP2* genes having a sequence derived from a particular source.

A naturally occurring 2μm-family plasmid is any plasmid having the features defined above as being essential features for a 2μm-family plasmid, which plasmid is found to naturally exist in yeast, i.e. has not been recombinantly modified to include heterologous sequence. Preferably the naturally occurring 2μm-family plasmid is selected from pSR1 (Accession No. X02398), pSB3 (Accession No. X02608) or pSB4 as obtained from *Zygosaccharomyces rouxii*, pSB1 or pSB2 (Accession No. NC_002055 or M18274) both as obtained from *Zygosaccharomyces bailli*, pSM1 (Accession No. NC_002054) as obtained from *Zygosaccharomyces fermentati*, pKD1 (Accession No. X03961) as obtained from *Kluyveromyces drosophilarum*, pPM1 as obtained from *Pichia membranaefaciens*, or, most preferably, the 2μm plasmid (Accession No. NC_001398 or J01347) as obtained from *Saccharomyces cerevisiae*. Accession numbers refer to deposits at the NCBI.

Preferably, other than the polynucleotide sequence insertion, deletion and/or substitution, the sequence of the inverted repeat adjacent to said *FLP* and/or *REP2* gene is derived from the sequence of the corresponding inverted repeat in the same naturally occurring 2μm-family plasmid as the sequence from which the gene is derived. Thus, for example, if the *FLP* gene is derived from the 2μm plasmid as obtained from *S. cerevisiae*, then it is preferred that the inverted repeat adjacent to the *FLP* gene has a sequence derived from the inverted repeat that is adjacent to the *FLP* gene in the 2μm plasmid as obtained from *S. cerevisiae*. If the *REP2* gene is derived from the 2μm plasmid as obtained from *S. cerevisiae*, then it is preferred that the inverted repeat adjacent to the *REP2* gene has a sequence derived from the inverted repeat that is adjacent to the *REP2* gene in the 2μm plasmid as obtained from *S. cerevisiae*.

Where, in the first preferred aspect of the invention, other than the polynucleotide sequence insertion, deletion and/or substitution, the *REP2* gene and the inverted repeat sequence have sequences derived from the corresponding regions of the 2μm plasmid as obtained from *S. cerevisiae*, then it is preferred that the polynucleotide sequence insertion, deletion and/or substitution occurs at a position between the first base of codon 59 of the *REP* gene and the last base before the FRT site in the adjacent inverted repeat, more preferably at a position between the first base of the inverted repeat and the last base before the FRT site, even more preferably at a position after the translation termination codon of the *REP2* gene and before the last base before the FRT site, such as at the first base after the end of the *REP2* coding sequence.

Where, other than the polynucleotide sequence insertion, deletion and/or substitution, the *REP2* gene and the inverted repeat sequence have sequences derived from the corresponding regions of the 2μm plasmid as obtained from *S. cerevisiae*, then in one embodiment, other than the polynucleotide sequence insertion, deletion and/or substitution, the sequence of the *REP2* gene and the

adjacent inverted repeat is as defined by SEQ ID NO:1 or variant thereof. In SEQ ID NO:1, the first base of codon 59 of the REP2 gene is represented by base number 175 and the last base before the FRT site is represented by base number 1216. The FRT sequence given here is the 55-base-pair sequence from Sadowski *et al*, 1986, pp7-10, *Mechanisms of Yeast Recombination* (Current Communications in Molecular Biology) CSHL. Ed. Klar, A. Strathern, J. N. In SEQ ID NO:1, the first base of the inverted repeat is represented by base number 887 and the first base after the translation termination codon of the *REP2* gene is represented by base number 892.

In an even more preferred embodiment of the first aspect of the invention, other than the polynucleotide sequence insertion, deletion and/or substitution, the *REP2* gene and the inverted repeat sequence have sequences derived from the corresponding regions of the 2μm plasmid as obtained from *S. cerevisiae* and, in the absence of the interruption the polynucleotide sequence insertion, deletion and/or substitution, comprise an *XcmI* site or an *FspI* site within the inverted repeat and the polynucleotide sequence insertion, deletion and/or substitution occurs at the *XcmI* site, or at the *FspI* site. In SEQ ID NO:1, the *XcmI* site is represented by base numbers 935-949 and the *FspI* site is represented by base numbers 1172-1177.

Where, in the second preferred aspect of the invention, other than the polynucleotide sequence insertion, deletion and/or substitution, the *FLP* gene and the adjacent inverted repeat sequence have sequences derived from the corresponding regions of the 2μm plasmid as obtained from *S. cerevisiae*, then it is preferred that the polynucleotide sequence insertion, deletion and/or substitution occurs at a position between the first base of codon 344 of the *FLP* gene and the last base before the FRT site, more preferably between the first base of the inverted repeat and the last base before the FRT site, yet more preferably between

the first base after the end of the FLP coding sequence and the last base before the FRT site, such as at the first base after the end of the FLP coding sequence.

Where, other than the polynucleotide sequence insertion, deletion and/or substitution, the *FLP* gene and the adjacent inverted repeat sequence have sequences derived from the corresponding regions of the 2μm plasmid as obtained from *S. cerevisiae*, then in one embodiment, other than the polynucleotide sequence insertion, deletion and/or substitution, the sequence of the *FLP* gene and the inverted repeat that follows the *FLP* gene is as defined by SEQ ID NO:2 or variant thereof. In SEQ ID NO:2, the first base of codon 344 of the *FLP* gene is represented by base number 1030 and the last base before the FRT site is represented by base number 1419, the first base of the inverted repeat is represented by base number 1090, and the first base after the end of the FLP coding sequence is represented by base number 1273.

In an even more preferred embodiment of the second preferred aspect of the invention, other than the polynucleotide sequence insertion, deletion and/or substitution, the *FLP* gene and the adjacent inverted repeat sequence have sequences derived from the corresponding regions of the 2μm plasmid as obtained from *S. cerevisiae* and, in the absence of the polynucleotide sequence insertion, deletion and/or substitution, comprise an *HgaI* site or an *FspI* site within the inverted repeat and the polynucleotide sequence insertion, deletion and/or substitution occurs at the cut formed by the action of *HgaI* on the *HgaI* site (*HgaI* cuts outside the 5bp sequence that it recognises), or at the *FspI*. In SEQ ID NO:2, the *HgaI* site is represented by base numbers 1262-1266 and the *FspI* site is represented by base numbers 1375-1380.

The skilled person will appreciate that the features of the plasmid defined by the first and second preferred aspects of the present invention are not mutually exclusive. Thus, a plasmid according to a third preferred aspect of the present

invention may comprise polynucleotide sequence insertions, deletions and/or substitutions between the first bases after the last functional codons of both of the *REP2* gene and the *FLP* gene and the last bases before the FRT sites in the inverted repeats adjacent to each of said genes, which polynucleotide sequence insertions, deletions and/or substitutions can be the same or different. For example, a plasmid according to a third aspect of the present invention may, other than the polynucleotide sequence insertions, deletions and/or substitutions, comprise the sequence of SEQ ID NO:1 or variant thereof and the sequence of SEQ ID NO:2 or variant thereof, each comprising a polynucleotide sequence insertion, deletion and/or substitution at a position as defined above for the first and second preferred aspects of the invention, respectively.

The skilled person will appreciate that the features of the plasmid defined by the first, second and third preferred aspects of the present invention do not exclude the possibility of the plasmid also having other sequence modifications. Thus, for example, a 2μm-family plasmid of the first, second and third preferred aspects of the present invention may additionally comprise a polynucleotide sequence insertion, deletion and/or substitution which is not at a position as defined above. Accordingly, the plasmid may additionally carry transgenes at a site other than the insertion sites of the invention.

Alternative insertion sites in 2μm plasmids are known in the art, but do not provide the advantages of using the insertion sites defined by the present invention. Nevertheless, plasmids which already include a polynucleotide sequence insertion, deletion and/or substitution at a site known in the art can be further modified by making one or more further modifications at one or more of the sites defined by the first, second and third preferred aspects of the present invention. The skilled person will appreciate that, as discussed in the introduction to this application, there are considerable technical limitations placed on the

insertion of transgenes at sites of 2μm-family plasmids other than as defined by the first and second aspects of the invention.

Typical modified 2μm plasmids known in the art include those described in Rose & Broach (1990, *Methods Enzymol.*, **185**, 234-279), such as plasmids pCV19, pCV20, CV_{neo}, which utilise an insertion at *Eco*RI in *FLP*, plasmids pCV21, pGT41 and pYE which utilise *Eco*RI in *D* as the insertion site, plasmid pHKB52 which utilises *Pst*I in *D* as the insertion site, plasmid pJDB248 which utilises an insertion at *Pst*I in *D* and *Eco*RI in *D*, plasmid pJDB219 in which *Pst*I in *D* and *Eco*RI in *FLP* are used as insertion sites, plasmid G18, plasmid pAB18 which utilises an insertion at *Cla*I in *FLP*, plasmids pGT39 and pA3, plasmids pYT11, pYT14 and pYT11-LEU which use *Pst*I in *D* as the insertion site, and plasmid PTY39 which uses *Eco*RI in *FLP* as the insertion site. Other 2μm plasmids include pSAC3, pSAC3U1, pSAC3U2, pSAC300, pSAC310, pSAC3C1, pSAC3PL1, pSAC3SL4, and pSAC3SC1 are described in EP 0 286 424 and Chinery & Hinchliffe (1989, *Curr. Genet.*, **16**, 21-25) which also described *Pst*I, *Eag*I or *Sna*BI as appropriate 2μm insertion sites. Further 2μm plasmids include pAYE255, pAYE316, pAYE443, pAYE522 (Kerry-Williams *et al*, 1998, *Yeast*, **14**, 161-169), pDB2244 (WO 00/44772) and pAYE329 (Sleep *et al*, 2001, *Yeast*, **18**, 403-421).

In one preferred embodiment, a 2μm-like plasmid as defined by the first, second and third preferred aspects of the present invention additionally comprises a polynucleotide sequence insertion, deletion and/or substitution which occurs within an untranscribed region around the ARS sequence. For example, in the 2μm plasmid obtained from *S. cerevisiae*, the untranscribed region around the ARS sequence extends from end of the D gene to the beginning of ARS sequence. Insertion into *Sna*BI (near the origin of replication sequence ARS) is described in Chinery & Hinchliffe, 1989, *Curr. Genet.*, **16**, 21-25. The skilled person will appreciate that an additional polynucleotide sequence insertion, deletion and/or

substitution can also occur within the untranscribed region at neighbouring positions to the *Sna*BI site described by Chinery & Hinchliffe.

A plasmid according to any of the first, second or third aspects of the present invention may be a plasmid capable of autonomous replication in yeast, such as a member of the *Saccharomyces*, *Kluyveromyces*, *Zygosaccharomyces*, or *Pichia* genus, such *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, *Kluyveromyces lactis*, *Pichia pastoris and Pichia membranaefaciens*, *Zygosaccharomyces rouxii*, *Zygosaccharomyces bailii*, *Zygosaccharomyces fermentati*, or *Kluyveromyces drosphilarum*. *S. cerevisiae* and *S. carlsbergensis* are thought to provide a suitable host cell for the autonomous replication of all known 2μm plasmids.

In a preferred embodiment, the, or at least one, polynucleotide sequence insertion, deletion and/or substitution included in a 2μm-family plasmid of the invention is a polynucleotide sequence insertion. Any polynucleotide sequence insertion may be used, so long as it is not unacceptably detrimental to the stability of the plasmid, by which we mean that the plasmid is at least 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40% 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9% or substantially 100% stable on non-selective media such as YEPD media compared to the unmodified plasmid, the latter of which is assigned a stability of 100%. Preferably, the above mentioned level of stability is seen after separately culturing yeast cells comprising the modified and unmodified plasmids in a culture medium for one, two, three, four, five, six, seven, eight, nine ten, 11, 12, 13, 14, 15, 16, 17, 18, 19 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more generations.

Where the plasmid comprises a selectable marker, higher levels of stability can be obtained when transformants are grown under selective conditions (e.g. in minimal

medium), since the medium can place a selective pressure on the host to retain the plasmid.

Stability in non-selective and selective (e.g. minimal) media can be determined using the methods set forth above. Stability in selective media can be demonstrated by the observation that the plasmids can be used to transform yeast to prototrophy.

Typically, the polynucleotide sequence insertion will be at least 4, 6, 8, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500 or more base pairs in length. Usually, the polynucleotide sequence insertion will be up to 1kb, 2kb, 3kb, 4kb, 5kb, 6kb, 7kb, 8kb, 9kb, 10kb or more in length. The skilled person will appreciate that the 2μm plasmid of the present invention may comprise multiple polynucleotide sequence insertions at different sites within the plasmid. Typically, the total length of polynucleotide sequence insertions is no more than 5kb, 10kb, 15kb, 20kb, 25kb or 30kb although greater total length insertion may be possible.

The polynucleotide sequence insertion may contain a transcribed region or may contain no transcribed region. A transcribed region may encode an open reading frame, or may be non-coding. The polynucleotide sequence insertion may contain both transcribed and non-transcribed regions.

A transcribed region is a region of DNA that can be transcribed by RNA polymerase, typically yeast RNA polymerase. A transcribed region can encode a functional RNA molecule, such as ribosomal or transfer RNA or an RNA molecule that can function as an antisense or RNA interference ("RNAi") molecule. Alternatively a transcribed region can encode a messenger RNA molecule (mRNA), which mRNA can contain an open reading frame (ORF) which can be translated *in vivo* to produce a protein. The term "protein" as used herein

includes all natural and non-natural proteins, polypeptides and peptides. Preferably, the ORF encodes a heterologous protein. By "heterologous protein" we mean a protein that is not naturally encoded by a 2µm-family plasmid. Preferably, therefore, the heterologous protein is not a FLP, REP1, REP2, or a RAF/D protein as encoded by any one of pSR1, pSB3 or pSB4 as obtained from *Z. rouxii*, pSB1 or pSB2 both as obtained from *Z. bailli*, pSM1 as obtained from *Z. fermentati*, pKD1 as obtained from *K. drosophilarum*, pPM1 as obtained from *P. membranaefaciens* and the 2µm plasmid as obtained from *S. cerevisiae*.

Where the polynucleotide sequence insertion encodes an open reading frame, then it may additionally comprise some polynucleotide sequence that does not encode an open reading frame (termed "non-coding region").

Non-coding region in the polynucleotide sequence insertion may contain one or more regulatory sequences, operatively linked to the open reading frame, which allow for the transcription of the open reading frame and/or translation of the resultant transcript.

The term "regulatory sequence" refers to a sequence that modulates (i.e., promotes or reduces) the expression (i.e., the transcription and/or translation) of an open reading frame to which it is operably linked. Regulatory regions typically include promoters, terminators, ribosome binding sites and the like. The skilled person will appreciate that the choice of regulatory region will depend upon the intended expression system. For example, promoters may be constitutive or inducible and may be cell- or tissue-type specific or non-specific.

Where the expression system is yeast, such as *Saccharomyces cerevisiae*, suitable promoters for *S. cerevisiae* include those associated with the *PGK1* gene, *GAL1* or *GAL10* genes, *TEF1*, *TEF2*, *PYK1*, *PMA1*, *CYC1*, *PHO5*, *TRP1*, *ADH1*, *ADH2*, the genes for glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate

decarboxylase, phosphofructokinase, triose phosphate isomerase, phosphoglucose isomerase, glucokinase, α-mating factor pheromone, a-mating factor pheromone, the *PRB1* promoter, the *PRA1* promoter, the *GPD1* promoter, and hybrid promoters involving hybrids of parts of 5' regulatory regions with parts of 5' regulatory regions of other promoters or with upstream activation sites (e.g. the promoter of EP-A-258 067).

Suitable transcription termination signals are well known in the art. Where the host cell is eukaryotic, the transcription termination signal is preferably derived from the 3' flanking sequence of a eukaryotic gene, which contains proper signals for transcription termination and polyadenylation. Suitable 3' flanking sequences may, for example, be those of the gene naturally linked to the expression control sequence used, i.e. may correspond to the promoter. Alternatively, they may be different. In that case, and where the host is a yeast, preferably *S. cerevisiae*, then the termination signal of the *S. cerevisiae ADH1*, *ADH2*, *CYC1*, or *PGK1* genes are preferred.

In one embodiment, the favoured regulatory sequences in yeast, such as *Saccharomyces cerevisiae*, include: a yeast promoter (e.g. the *Saccharomyces cerevisiae PRB1* promoter), as taught in EP 431 880; and a transcription terminator, preferably the terminator from *Saccharomyces ADH1*, as taught in EP 60 057.

It may be beneficial for the non-coding region to incorporate more than one DNA sequence encoding a translational stop codon, such as UAA, UAG or UGA, in order to minimise translational read-through and thus avoid the production of elongated, non-natural fusion proteins. The translation stop codon UAA is preferred. Preferably, at least two translation stop codons are incorporated.

The term "operably linked" includes within its meaning that a regulatory sequence is positioned within any non-coding region such that it forms a relationship with an open reading frame that permits the regulatory region to exert an effect on the open reading frame in its intended manner. Thus a regulatory region "operably linked" to an open reading frame is positioned in such a way that the regulatory region is able to influence transcription and/or translation of the open reading frame in the intended manner, under conditions compatible with the regulatory sequence.

Where the polynucleotide sequence insertion as defined by the first, second or third aspects of the present invention includes an open reading frame that encodes a protein, then it may be advantageous for the encoded protein to be secreted. In that case, a sequence encoding a secretion leader sequence which, for example, comprises most of the natural HSA secretion leader, plus a small portion of the *S. cerevisiae* α-mating factor secretion leader as taught in WO 90/01063, may be included in the open reading frame.

Alternatively, the encoded protein may be intracellular.

In one preferred embodiment, at least one polynucleotide sequence insertion as defined by the first, second or third aspects of the present invention includes an open reading frame comprising a sequence that encodes a yeast protein. In another preferred embodiment, at least one polynucleotide sequence insertion as defined by the first, second or third aspects of the present invention includes an open reading frame comprising a sequence that encodes a yeast protein from the same host from which the 2μm-like plasmid is derived.

In another preferred embodiment, at least one polynucleotide sequence insertion as defined by the first, second or third aspects of the present invention includes an open reading frame comprising a sequence that encodes a protein involved in

protein folding, or which has chaperone activity or is involved in the unfolded protein response (Stanford Genome Database (SGD), http://db.yeastgenome.org). Preferred proteins include *AHA1*, *CCT2*, *CCT3*, *CCT4*, *CCT5*, *CCT6*, *CCT7*, *CCT8*, *CNS1*, *CPR3*, *CPR6*, *ERO1*, *EUG1*, *FMO1*, *HCH1*, *HSP10*, *HSP12*, *HSP104*, *HSP26*, *HSP30*, *HSP42*, *HSP60*, *HSP78*, *HSP82*, *JEM1*, *MDJ1*, *MDJ2*, *MPD1*, *MPD2*, *PDI1*, *PFD1*, *ABC1*, *APJ1*, *ATP11*, *ATP12*, *BTT1*, *CDC37*, *CNS1*, *CPR6*, *CPR7*, *HSC82*, *KAR2*, *LHS1*, *MGE1*, *MRS11*, *NOB1*, *ECM10*, *SSA1*, *SSA2*, *SSA3*, *SSA4*, *SSC1*, *SSE2*, *SIL1*, *SLS1*, *ORM1*, *UBI4*, *ORM2*, *PER1*, *PTC2*, *PSE1* and HAC1 or a truncated intronless *HAC1* (Valkonen *et al.* 2003, *Applied Environ. Micro.* 69, 2065).

A preferred chaperone is protein disulphide isomerase (PDI) or a fragment or variant thereof having an equivalent ability to catalyse the formation of disulphide bonds within the lumen of the endoplasmic reticulum (ER). By "PDI" we include any protein having the ability to reactivate the ribonuclease activity against RNA of scrambled ribonuclease as described in EP 0 746 611 and Hillson *et al*, 1984, *Methods Enzymol.*. **107**, 281-292.

Protein disulphide isomerase is an enzyme which typically catalyzes thiol:disulphide interchange reactions, and is a major resident protein component of the E.R. lumen in secretory cells. A body of evidence suggests that it plays a role in secretory protein biosynthesis (Freedman, 1984, *Trends Biochem. Sci.*, 9, 438-41) and this is supported by direct cross-linking studies *in situ* (Roth and Pierce, 1987, *Biochemistry*, 26, 4179-82). The finding that microsomal membranes deficient in PDI show a specific defect in cotranslational protein disulphide formation (Bulleid and Freedman, 1988, *Nature*, 335, 649-51) implies that the enzyme functions as a catalyst of native disulphide bond formation during the biosynthesis of secretory and cell surface proteins. This role is consistent with what is known of the enzyme's catalytic properties *in vitro*, it catalyzes thiol: disulphide interchange reactions leading to net protein disulphide formation,

breakage or isomerization, and can typically catalyze protein folding and the formation of native disulphide bonds in a wide variety of reduced, unfolded protein substrates (Freedman *et al.*, 1989, *Biochem. Soc. Symp.*, **55**, 167-192). The DNA and amino acid sequence of the enzyme is known for several species (Scherens *et al*, 1991, *Yeast*, **7**, 185-193; Farquhar *et al*, 1991, *Gene*, **108**, 81-89; EP074661; EP0293793; EP0509841) and there is increasing information on the mechanism of action of the enzyme purified to homogeneity from mammalian liver (Creighton *et al*, 1980, *J. Mol. Biol.*, **142**, 43-62; Freedman *et al*, 1988, *Biochem. Soc. Trans.*, **16**, 96-9; Gilbert, 1989, *Biochemistry*, **28**, 7298-7305; Lundstrom and Holmgren, 1990, *J. Biol. Chem.*, **265**, 9114-9120; Hawkins and Freedman, 1990, *Biochem. J.*, **275**, 335-339). Of the many protein factors currently implicated as mediators of protein folding, assembly and translocation in the cell (Rothman, 1989, *Cell*, **59**, 591-601), PDI has a well-defined catalytic activity.

PDI is readily isolated from mammalian tissues and the homogeneous enzyme is a homodimer (2×57 kD) with characteristically acidic pI (4.0-4.5) (Hillson *et al*, 1984, *Methods Enzymol.*, **107**, 281-292). The enzyme has also been purified from wheat and from the alga *Chlamydomonas reinhardii* (Kaska *et al*, 1990, *Biochem. J.*, **268**, 63-68), rat (Edman *et al*, 1985, *Nature*, **317**, 267-270), bovine (Yamauchi *et al*, 1987, *Biochem. Biophys. Res. Comm.*, **146**, 1485-1492), human (Pihlajaniemi *et al*, 1987, *EMBO J.*, **6**, 643-9), yeast (Scherens *et al*, *supra*; Farquhar *et al*, *supra*) and chick (Parkkonen *et al*, 1988, *Biochem. J.*, **256**, 1005-1011). The proteins from these vertebrate species show a high degree of sequence conservation throughout and all show several overall features first noted in the rat PDI sequence (Edman *et al.*, 1985, *op. cit.*).

A yeast protein disulphide isomerase precursor, PDI1, can be found as Genbank accession no. CAA42373 or BAA00723. It has the following sequence of 522 amino acids:

```
1   mkfsagavls wssllllassv faqqeavape dsavvklatd sfneyiqshd lvlaeffapw
61  cghcknmape yvkaaetlve knitlaqidc tenqdlcmeh nipgfpslki fknsdvnnsi
121 dyegprtaea ivqfmikqsq pavavvadlp aylanetfvt pvivqsgkid adfnatfysm
181 ankhfndydf vsaenadddf klsiylpsam depvvyngkk adiadadvfe kwlqvealpy
241 fgeidgsvfa qyvesglplg ylfyndeeel eeykplftel akknrglmnf vsidarkfgr
301 hagnlnmkeq fplfaihdmt edlkyglpql seeafdelsd kivleskaie slvkdflkgd
361 aspivksqei fenqdssvfq lvgknhdeiv ndpkkdvlvl yyapwcghck rlaptyqela
421 dtyanatsdv liakldhten dvrgvviegy ptivlypggk ksesvvyqgs rsldslfdfi
481 kenghfdvdg kalyeeaqek aaeeadadae ladeedaihd el
```

An alternative PDI sequence can be found as Genbank accession no. CAA38402. It has the following sequence of 530 amino acids

```
1   mkfsagavls wssllllassv faqqeavape dsavvklatd sfneyiqshd lvlaeffapw
61  cghcknmape yvkaaetlve knitlaqidc tenqdlcmeh nipgfpslki fknrdvnnsi
121 dyegprtaea ivqfmikqsq pavavvadlp aylanetfvt pvivqsgkid adfnatfysm
181 ankhfndydf vsaenadddf klsiylpsam depvvyngkk adiadadvfe kwlqvealpy
241 fgeidgsvfa qyvesglplg ylfyndeeel eeykplftel akknrglmnf vsidarkfgr
301 hagnlnmkeq fplfaihdmt edlkyglpql seeafdelsd kivleskaie slvkdflkgd
361 aspivksqei fenqdssvfq lvgknhdeiv ndpkkdvlvl yyapwcghck rlaptyqela
421 dtyanatsdv liakldhten dvrgvviegy ptivlypggk ksesvvyqgs rsldslfdfi
481 kenghfdvdg kalyeeaqek aaeeaeadae aeadadaela deedaihdel
```

Variants and fragments of the above PDI sequences, and variants of other naturally occurring PDI sequences are also included in the present invention. A "variant", in the context of PDI, refers to a protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in a protein whose basic properties, for example enzymatic activity (type of and specific activity), thermostability, activity in a certain pH-range (pH-stability) have not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein.

By "conservative substitutions" is intended combinations such as Val, Ile, Leu, Ala, Met; Asp, Glu; Asn, Gln; Ser, Thr, Gly, Ala; Lys, Arg, His; and Phe, Tyr, Trp. Preferred conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr.

A "variant" typically has at least 25%, at least 50%, at least 60% or at least 70%, preferably at least 80%, more preferably at least 90%, even more preferably at least 95%, yet more preferably at least 99%, most preferably at least 99.5% sequence identity to the polypeptide from which it is derived.

The percent sequence identity between two polypeptides may be determined using suitable computer programs, as discussed below. Such variants may be natural or made using the methods of protein engineering and site-directed mutagenesis as are well known in the art.

A "fragment", in the context of PDI, refers to a protein wherein at one or more positions there have been deletions. Thus the fragment may comprise at most 5, 10, 20, 30, 40 or 50% of the complete sequence of the full mature PDI protein. Typically a fragment comprises up to 60%, more typically up to 70%, preferably up to 80%, more preferably up to 90%, even more preferably up to 95%, yet more preferably up to 99% of the complete sequence of the full PDI protein. Particularly preferred fragments of PDI protein comprise one or more whole domains of the desired protein.

It is particularly preferred that a plasmid according to a first, second or third aspects of the invention includes, either within a polynucleotide sequence insertion, or elsewhere on the plasmid, an open reading frame encoding a protein comprising the sequence of albumin or a fragment or variant thereof. Alternatively, the host cell into which the plasmid is transformed may include

within its genome a polynucleotide sequence encoding a protein comprising the sequence of albumin or a fragment or variant thereof, either as an endogenous or heterologous sequence.

By "albumin" we include a protein having the sequence of an albumin protein obtained from any source. Typically the source is mammalian. In one preferred embodiment the serum albumin is human serum albumin ("HSA"). The term "human serum albumin" includes the meaning of a serum albumin having an amino acid sequence naturally occurring in humans, and variants thereof. Preferably the albumin has the amino acid sequence disclosed in WO 90/13653 or a variant thereof. The HSA coding sequence is obtainable by known methods for isolating cDNA corresponding to human genes, and is also disclosed in, for example, EP 73 646 and EP 286 424.

In another preferred embodiment the "albumin" has the sequence of bovine serum albumin. The term "bovine serum albumin" includes the meaning of a serum albumin having an amino acid sequence naturally occurring in cows, for example as taken from Swissprot accession number P02769, and variants thereof as defined below. The term "bovine serum albumin" also includes the meaning of fragments of full-length bovine serum albumin or variants thereof, as defined below.

In another preferred embodiment the albumin is an albumin derived from (i.e. has the sequence of) one of serum albumin from dog (e.g. see Swissprot accession number P49822), pig (e.g. see Swissprot accession number P08835), goat (e.g. as available from Sigma as product no. A2514 or A4164), turkey (e.g. see Swissprot accession number O73860), baboon (e.g. as available from Sigma as product no. A1516), cat (e.g. see Swissprot accession number P49064), chicken (e.g. see Swissprot accession number P19121), ovalbumin (e.g. chicken ovalbumin) (e.g. see Swissprot accession number P01012), donkey (e.g. see Swissprot accession number P39090), guinea pig (e.g. as available from Sigma as product no. A3060,

A2639, O5483 or A6539), hamster (e.g. as available from Sigma as product no. A5409), horse (e.g. see Swissprot accession number P35747), rhesus monkey (e.g. see Swissprot accession number Q28522), mouse (e.g. see Swissprot accession number O89020), pigeon (e.g. as defined by Khan *et al*, 2002, *Int. J. Biol. Macromol.*, 30(3-4),171-8), rabbit (e.g. see Swissprot accession number P49065), rat (e.g. see Swissprot accession number P36953) and sheep (e.g. see Swissprot accession number P14639) and includes variants and fragments thereof as defined below.

Many naturally occurring mutant forms of albumin are known. Many are described in Peters, (1996, *All About Albumin: Biochemistry, Genetics and Medical Applications*, Academic Press, Inc., San Diego, California, p.170-181). A variant as defined above may be one of these naturally occurring mutants.

A "variant albumin" refers to an albumin protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in an albumin protein for which at least one basic property, for example binding activity (type of and specific activity e.g. binding to bilirubin), osmolarity (oncotic pressure, colloid osmotic pressure), behaviour in a certain pH-range (pH-stability) has not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein.

By "conservative substitutions" is intended combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. Such variants may be made by techniques well known in the art, such as by site-directed mutagenesis as disclosed in US Patent No 4,302,386 issued 24 November 1981 to Stevens, incorporated herein by reference.

Typically an albumin variant will have more than 40%, usually at least 50%, more typically at least 60%, preferably at least 70%, more preferably at least 80%, yet more preferably at least 90%, even more preferably at least 95%, most preferably at least 98% or more sequence identity with naturally occurring albumin. The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally. The alignment may alternatively be carried out using the Clustal W program (Thompson *et al.*, 1994). The parameters used may be as follows:

Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method: x percent. Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05. Scoring matrix: BLOSUM.

The term "fragment" as used above includes any fragment of full-length albumin or a variant thereof, so long as at least one basic property, for example binding activity (type of and specific activity e.g. binding to bilirubin), osmolarity (oncotic pressure, colloid osmotic pressure), behaviour in a certain pH-range (pH-stability) has not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein. A fragment will typically be at least 50 amino acids long. A fragment may comprise at least one whole sub-domain of albumin. Domains of HSA have been expressed as recombinant proteins (Dockal, M. *et al.*, 1999, *J. Biol. Chem.*, 274, 29303-29310), where domain I was defined as consisting of amino acids 1-197, domain II was defined as consisting of amino acids 189-385 and domain III was defined as consisting of amino acids 381-585. Partial overlap of the domains occurs because of the extended α-helix structure (h10-h1) which exists between domains I and II,

and between domains II and III (Peters, 1996, *op. cit.*, Table 2-4). HSA also comprises six sub-domains (sub-domains IA, IB, IIA, IIB, IIIA and IIIB). Sub-domain IA comprises amino acids 6-105, sub-domain IB comprises amino acids 120-177, sub-domain IIA comprises amino acids 200-291, sub-domain IIB comprises amino acids 316-369, sub-domain IIIA comprises amino acids 392-491 and sub-domain IIIB comprises amino acids 512-583. A fragment may comprise a whole or part of one or more domains or sub-domains as defined above, or any combination of those domains and/or sub-domains.

Thus the polynucleotide insertion may comprise an open reading frame that encodes albumin or a variant or fragment thereof.

The skilled person will also appreciate that the open reading frame of any other gene or variant, or part or either, can be utilised to form a whole or part of an open reading frame in forming a polynucleotide sequence insertion for use with the present invention. For example, the open reading frame may encode a protein comprising any sequence, be it a natural protein (including a zymogen), or a variant, or a fragment (which may, for example, be a domain) of a natural protein; or a totally synthetic protein; or a single or multiple fusion of different proteins (natural or synthetic). Such proteins can be taken, but not exclusively, from the lists provided in WO 01/79258, WO 01/79271, WO 01/79442, WO 01/79443, WO 01/79444 and WO 01/79480, or a variant or fragment thereof; the disclosures of which are incorporated herein by reference. Although these patent applications present the list of proteins in the context of fusion partners for albumin, the present invention is not so limited and, for the purposes of the present invention, any of the proteins listed therein may be presented alone or as fusion partners for albumin, the Fc region of immunoglobulin, transferrin, lactoferrin or any other protein or fragment or variant of any of the above, including fusion proteins comprising any of the above, as a desired polypeptide. Examples of transferrin fusions are given in US patent applications US2003/0221201 and

US2003/0226155, Shin, *et al.*, 1995, *Proc Natl Acad Sci U S A*, 92, 2820; Ali, *et al.*, 1999, *J Biol Chem*, 274, 24066; Mason, *et al.*, 2002, *Biochemistry*, 41, 9448.

Preferred other examples of desirable proteins for expression by the present invention includes sequences comprising the sequence of a monoclonal antibody, an etoposide, a serum protein (such as a blood clotting factor), antistasin, a tick anticoagulant peptide, transferrin, lactoferrin, endostatin, angiostatin, collagens, immunoglobulins, Fab' fragments, F(ab')2, scAb, scFv, interferons, interleukins, IL10, IL11, IL2, interferon α species and sub-species, interferon β species and sub-species, interferon γ species and sub-species, leptin, CNTF, $CNTF_{Ax15}$, IL1-receptor antagonist, erythropoetin (EPO) and EPO mimics, thrombopoetin (TPO) and TPO mimics, prosaptide, cyanovirin-N, 5-helix, T20 peptide, T1249 peptide, HIV gp41, HIV gp120, urokinase, prourokinase, tPA (tissue plasminogen activator), hirudin, platelet derived growth factor, parathyroid hormone, proinsulin, insulin, glucagon, glucagon-like peptides, insulin-like growth factor, calcitonin, growth hormone, transforming growth factor β, tumour necrosis factor, G-CSF, GM-CSF, M-CSF, FGF, coagulation factors in both pre and active forms, including but not limited to plasminogen, fibrinogen, thrombin, pre-thrombin, pro-thrombin, von Willebrand's factor, α_1-antitrypsin, plasminogen activators, Factor VII, Factor VIII, Factor IX, Factor X and Factor XIII, nerve growth factor, LACI (lipoprotein associated coagulation inhibitor, also known as tissue factor pathway inhibitor or extrinsic pathway inhibitor), platelet-derived endothelial cell growth factor (PD-ECGF), glucose oxidase, serum cholinesterase, aprotinin, amyloid precursor, inter-alpha trypsin inhibitor, antithrombin III, apo-lipoprotein species, Protein C, Protein S, a variant or fragment or fusion protein of any of the above.

A "variant", in the context of the above-listed proteins, refers to a protein wherein at one or more positions there have been amino acid insertions, deletions, or substitutions, either conservative or non-conservative, provided that such changes result in a protein whose basic properties, for example enzymatic activity or receptor

binding (type of and specific activity), thermostability, activity in a certain pH-range (pH-stability) have not significantly been changed. "Significantly" in this context means that one skilled in the art would say that the properties of the variant may still be different but would not be unobvious over the ones of the original protein.

By "conservative substitutions" is intended combinations such as Val, Ile, Leu, Ala, Met; Asp, Glu; Asn, Gln; Ser, Thr, Gly, Ala; Lys, Arg, His; and Phe, Tyr, Trp. Preferred conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr.

A "variant" typically has at least 25%, at least 50%, at least 60% or at least 70%, preferably at least 80%, more preferably at least 90%, even more preferably at least 95%, yet more preferably at least 99%, most preferably at least 99.5% sequence identity to the polypeptide from which it is derived.

The percent sequence identity between two polypeptides may be determined using suitable computer programs, for example the GAP program of the University of Wisconsin Genetic Computing Group and it will be appreciated that percent identity is calculated in relation to polypeptides whose sequence has been aligned optimally.

The alignment may alternatively be carried out using the Clustal W program (Thompson *et al.*, (1994) *Nucleic Acids Res.*, 22(22), 4673-80). The parameters used may be as follows:

- Fast pairwise alignment parameters: K-tuple(word) size; 1, window size; 5, gap penalty; 3, number of top diagonals; 5. Scoring method; x percent.
- Multiple alignment parameters: gap open penalty; 10, gap extension penalty; 0.05.
- Scoring matrix: BLOSUM.

Such variants may be natural or made using the methods of protein engineering and site-directed mutagenesis as are well known in the art.

A "fragment", in the context of the above-listed proteins, refers to a protein wherein at one or more positions there have been deletions. Thus the fragment may comprise at most 5, 10, 20, 30, 40 or 50% of the complete sequence of the full mature polypeptide. Typically a fragment comprises up to 60%, more typically up to 70%, preferably up to 80%, more preferably up to 90%, even more preferably up to 95%, yet more preferably up to 99% of the complete sequence of the full desired protein. Particularly preferred fragments of a desired protein comprise one or more whole domains of the desired protein.

It is particularly preferred that a plasmid according to a first, second or third aspects of the invention includes, either within a polynucleotide sequence insertion, or elsewhere on the plasmid, an open reading frame encoding a protein comprising the sequence of albumin or a fragment or variant thereof, or any other protein take from the examples above (fused or unfused to a fusion partner) and at least one other heterologous sequence, wherein the at least one other heterologous sequence may contain a transcribed region, such as an open reading frame. In one embodiment, the open reading frame may encode a protein comprising the sequence of a yeast protein. In another embodiment the open reading frame may encode a protein comprising the sequence of a protein involved in protein folding, or which has chaperone activity or is involved in the unfolded protein response, preferably protein disulphide isomerase.

The present invention also provides a method of preparing a plasmid of the invention, which method comprises –

(a) providing a 2μm-family plasmid comprising a *REP2* gene or an *FLP* gene and an inverted repeat adjacent to said gene;

(b) providing a polynucleotide sequence and inserting the polynucleotide sequence into the plasmid at a position according to the first, second or third preferred aspects of the invention; and/or

(c) additionally or as an alternative to step (b), deleting some or all of the nucleotide bases at the positions according to the first, second or third preferred aspects of the invention; and/or

(d) additionally or as an alternative to either of steps (b) and (c), substituting some or all of the nucleotide bases at the positions according to the first, second or third preferred aspects of the invention with alternative nucleotide bases.

Steps (b), (c) and (d) can be achieved using techniques well known in the art, including cloning techniques, site-directed mutagenesis and the like, such as are described in by Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2001, 3rd edition, the contents of which are incorporated herein by reference. For example, one such method involves ligation via cohesive ends. Compatible cohesive ends can be generated on a DNA fragment for insertion and plasmid by the action of suitable restriction enzymes. These ends will rapidly anneal through complementary base pairing and remaining nicks can be closed by the action of DNA ligase.

A further method uses synthetic double stranded oligonucleotide linkers and adaptors. DNA fragments with blunt ends are generated by bacteriophage T4 DNA polymerase or *E.coli* DNA polymerase I which remove protruding 3' termini and fill in recessed 3' ends. Synthetic linkers and pieces of blunt-ended double-stranded DNA, which contain recognition sequences for defined restriction enzymes, can be ligated to blunt-ended DNA fragments by T4 DNA ligase. They are subsequently

digested with appropriate restriction enzymes to create cohesive ends and ligated to an expression vector with compatible termini. Adaptors are also chemically synthesised DNA fragments which contain one blunt end used for ligation but which also possess one preformed cohesive end. Alternatively a DNA fragment or DNA fragments can be ligated together by the action of DNA ligase in the presence or absence of one or more synthetic double stranded oligonucleotides optionally containing cohesive ends.

Synthetic linkers containing a variety of restriction endonuclease sites are commercially available from a number of sources including Sigma-Genosys Ltd, London Road, Pampisford, Cambridge, United Kingdom.

Accordingly, the present invention also provides a plasmid obtainable by the above method.

The present invention also provides a host cell comprising a plasmid as defined above. The host cell may be any type of cell. Bacterial and yeast host cells are preferred. Bacterial host cells may be useful for cloning purposes. Yeast host cells may be useful for expression of genes present in the plasmid.

In one embodiment the host cell is a cell in which the plasmid is stable as a multicopy plasmid. Plasmids obtained from one yeast type can be maintained in other yeast types (Irie *et al*, 1991, *Gene*, **108**(1), 139-144; Irie *et al*, 1991, *Mol. Gen. Genet.*, 225(2), 257-265). For example, pSR1 from *Zygosaccharomyces rouxii* can be maintained in *Saccharomyces cerevisiae*. Where the plasmid is based on pSR1, pSB3 or pSB4 the host cell may be *Zygosaccharomyces rouxii*, where the plasmid is based on pSB1 or pSB2 the host cell may be *Zygosaccharomyces bailli*, where the plasmid is based on pPM1 the host cell may be *Pichia membranaefaciens*, where the plasmid is based on pSM1 the host cell may be *Zygosaccharomyces fermentati*, where the plasmid is based on pKD1 the

host cell may be *Kluyveromyces drosophilarum* and where the plasmid is based on the 2μm plasmid the host cell may be *Saccharomyces cerevisiae* or *Saccharomyces carlsbergensis*. A 2μm-family plasmid of the invention can be said to be "based on" a naturally occurring plasmid if it comprises one, two or preferably three of the genes *FLP*, *REP1* and *REP2* having sequences derived from that naturally occurring plasmid.

A plasmid as defined above, may be introduced into a host through standard techniques. With regard to transformation of prokaryotic host cells, see, for example, Cohen *et al* (1972) *Proc. Natl. Acad. Sci. USA* 69, 2110 and Sambrook *et al* (2001) *Molecular Cloning, A Laboratory Manual*, 3rd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. Transformation of yeast cells is described in Sherman *et al* (1986) *Methods In Yeast Genetics, A Laboratory Manual*, Cold Spring Harbor, NY. The method of Beggs (1978) *Nature* 275, 104-109 is also useful. Methods for the transformation of *S. cerevisiae* are taught generally in EP 251 744, EP 258 067 and WO 90/01063, all of which are incorporated herein by reference. With regard to vertebrate cells, reagents useful in transfecting such cells, for example calcium phosphate and DEAE-dextran or liposome formulations, are available from Stratagene Cloning Systems, or Life Technologies Inc., Gaithersburg, MD 20877, USA.

Electroporation is also useful for transforming cells and is well known in the art for transforming yeast cell, bacterial cells and vertebrate cells. Methods for transformation of yeast by electroporation are disclosed in Becker & Guarente (1990) *Methods Enzymol.* 194, 182.

Generally, the plasmid will transform not all of the hosts and it will therefore be necessary to select for transformed host cells. Thus, a plasmid according to any one of the first, second or third aspects of the present invention may comprise a selectable marker, either within a polynucleotide sequence insertion, or elsewhere

on the plasmid, including but not limited to bacterial selectable marker and/or a yeast selectable marker. A typical bacterial selectable marker is the β-lactamase gene although many others are known in the art. Typical yeast selectable marker include *LEU2* (or an equivalent gene encoding a protein with the activity of β-lactamase malate dehydrogenase), *TRP1, HIS3, HIS4, URA3, URA5, SFA1, ADE2, MET15, LYS5, LYS2, ILV2, FBA1* and *PGK1*. Those skilled in the art will appreciate that any gene whose chromosomal deletion or inactivation results in an inviable host, so called essential genes, can be used as a selective marker if a functional gene is provided on the plasmid, as demonstrated for *PGK1* in a *pgk1* yeast strain (Piper and Curran, 1990, *Curr. Genet.* 17, 119). Suitable essential genes can be found within the Stanford Genome Database (SGD), http:://db.yeastgenome.org).

Additionally, a plasmid according to any one of the first, second or third aspects of the present invention may comprise more than one selectable marker, either within a polynucleotide sequence insertion, or elsewhere on the plasmid.

One selection technique involves incorporating into the expression vector a DNA sequence marker, with any necessary control elements, that codes for a selectable trait in the transformed cell. These markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture, and tetracyclin, kanamycin or ampicillin (i.e. β-lactamase) resistance genes for culturing in *E.coli* and other bacteria. Alternatively, the gene for such selectable trait can be on another vector, which is used to co-transform the desired host cell.

Another method of identifying successfully transformed cells involves growing the cells resulting from the introduction of a plasmid of the invention, optionally to allow the expression of a recombinant polypeptide (i.e. a polypeptide which is encoded by a polynucleotide sequence on the plasmid and is heterologous to the host cell, in the sense that that polypeptide is not naturally produced by the host). Cells can be

harvested and lysed and their DNA or RNA content examined for the presence of the recombinant sequence using a method such as that described by Southern (1975) *J. Mol. Biol.* **98**, 503 or Berent *et al* (1985) *Biotech.* 3, 208, or other methods of DNA and RNA analysis common in the art. Alternatively, the presence of a polypeptide in the supernatant of a culture of a transformed cell can be detected using antibodies.

In addition to directly assaying for the presence of recombinant DNA, successful transformation can be confirmed by well known immunological methods when the recombinant DNA is capable of directing the expression of the protein. For example, cells successfully transformed with an expression vector produce proteins displaying appropriate antigenicity. Samples of cells suspected of being transformed are harvested and assayed for the protein using suitable antibodies.

Thus, in addition to the transformed host cells themselves, the present invention also contemplates a culture of those cells, preferably a monoclonal (clonally homogeneous) culture, or a culture derived from a monoclonal culture, in a nutrient medium.

Transformed host cells may then be cultured for a sufficient time and under appropriate conditions known to those skilled in the art; and in view of the teachings disclosed herein, to permit the expression of any ORF(s) in the one or more polynucleotide sequence insertions within the plasmid.

The present invention thus also provides a method for producing a protein comprising the steps of (a) providing a plasmid according to the first, second or third aspects of the invention as defined above; (b) providing a suitable host cell; (c) transforming the host cell with the plasmid; and (d) culturing the transformed host cell in a culture medium, thereby to produce the protein.

Many expression systems are known, including bacteria (for example *E. coli* and *Bacillus subtilis*), yeasts, filamentous fungi (for example *Aspergillus*), plant cells, whole plants, animal cells and insect cells.

In one embodiment the preferred host cells are the yeasts in which the plasmid is capable of being maintained as a stable multicopy plasmid. Such yeasts include *Saccharomyces cerevisiae, Kluyveromyces lactis, Pichia pastoris, Zygosaccharomyces rouxii, Zygosaccharomyces bailli, Zygos-accharomyces fermentati,* and *Kluyveromyces drosophilarum.*

A plasmid is capable of being maintained as a stable multicopy plasmid in a host, if the plasmid contains, or is modified to contain, a selectable (e.g. *LEU2*) marker, and stability, as measured by the loss of the marker, is at least 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9% or substantially 100% after one, two, three, four, five, six, seven, eight, nine ten, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more generations. Loss of a marker can be assessed as described above, with reference to Chinery & Hinchliffe (1989, *Curr. Genet.*, 16, 21-25).

It is particularly advantageous to use a yeast deficient in one or more protein mannosyl transferases involved in O-glycosylation of proteins, for instance by disruption of the gene coding sequence.

Recombinantly expressed proteins can be subject to undesirable post-translational modifications by the producing host cell. For example, the albumin protein sequence does not contain any sites for N-linked glycosylation and has not been reported to be modified, in nature, by O-linked glycosylation. However, it has been found that recombinant human albumin ("rHA") produced in a number of yeast species can be modified by O-linked glycosylation, generally involving

mannose. The mannosylated albumin is able to bind to the lectin Concanavalin A. The amount of mannosylated albumin produced by the yeast can be reduced by using a yeast strain deficient in one or more of the *PMT* genes (WO 94/04687). The most convenient way of achieving this is to create a yeast which has a defect in its genome such that a reduced level of one of the Pmt proteins is produced. For example, there may be a deletion, insertion or transposition in the coding sequence or the regulatory regions (or in another gene regulating the expression of one of the *PMT* genes) such that little or no Pmt protein is produced. Alternatively, the yeast could be transformed to produce an anti-Pmt agent, such as an anti-Pmt antibody.

If a yeast other than *S. cerevisiae* is used, disruption of one or more of the genes equivalent to the *PMT* genes of *S. cerevisiae* is also beneficial, e.g. in *Pichia pastoris* or *Kluyveromyces lactis*. The sequence of *PMT1* (or any other *PMT* gene) isolated from *S. cerevisiae* may be used for the identification or disruption of genes encoding similar enzymatic activities in other fungal species. The cloning of the *PMT1* homologue of *Kluyveromyces lactis* is described in WO 94/04687.

The yeast will advantageously have a deletion of the *HSP150* and/or *YAP3* genes as taught respectively in WO 95/33833 and WO 95/23857.

The present application also provides a method of producing a protein comprising the steps of providing a host cell as defined above, which host cell comprises a plasmid of the present invention and culturing the host cell in a culture medium thereby to produce the protein. The culture medium may be non-selective or place a selective pressure on the stable multicopy maintenance of the plasmid.

A method of producing a protein expressed from a plasmid of the invention preferably further comprise the step of isolating the thus produced protein from the cultured host cell or the culture medium.

The thus produced protein may be present intracellularly or, if secreted, in the culture medium and/or periplasmic space of the host cell. The protein may be isolated from the cell and/or culture medium by many methods known in the art. For example purification techniques for the recovery of recombinantly expressed albumin have been disclosed in: WO 92/04367, removal of matrix-derived dye; EP 464 590, removal of yeast-derived colorants; EP 319 067, alkaline precipitation and subsequent application of the albumin to a lipophilic phase; and WO 96/37515, US 5 728 553 and WO 00/44772, which describe complete purification processes; all of which are incorporated herein by reference. Proteins other than albumin may be purified from the culture medium by any technique that has been found to be useful for purifying such proteins.

Such well-known methods include ammonium sulphate or ethanol precipitation, acid or solvent extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, concentration, dilution, pH adjustment, diafiltration, ultrafiltration, high performance liquid chromatography ("HPLC"), reverse phase HPLC, conductivity adjustment and the like.

In one embodiment, any one or more of the above mentioned techniques may be used to further purifying the thus isolated protein to a commercially acceptable level of purity. By commercially acceptable level of purity, we include the provision of the protein at a concentration of at least 0.01 $g.L^{-1}$, 0.02 $g.L^{-1}$, 0.03 $g.L^{-1}$, 0.04 $g.L^{-1}$, 0.05 $g.L^{-1}$,0.06 $g.L^{-1}$,0.07 $g.L^{-1}$, 0.08 $g.L^{-1}$, 0.09 $g.L^{-1}$, 0.1 $g.L^{-1}$, 0.2 $g.L^{-1}$, 0.3 $g.L^{-1}$, 0.4 $g.L^{-1}$, 0.5 $g.L^{-1}$, 0.6 $g.L^{-1}$, 0.7 $g.L^{-1}$, 0.8 $g.L^{-1}$, 0.9 $g.L^{-1}$, 1 $g.L^{-1}$, 2 $g.L^{-1}$, 3 $g.L^{-1}$, 4 $g.L^{-1}$, 5 $g.L^{-1}$, 6 $g.L^{-1}$, 7 $g.L^{-1}$, 8 $g.L^{-1}$, 9 $g.L^{-1}$, 10 $g.L^{-1}$, 15 $g.L^{-1}$, 20 $g.L^{-1}$, 25

$g.L^{-1}$, 30 $g.L^{-1}$, 40 $g.L^{-1}$,50 $g.L^{-1}$, 60 $g.L^{-1}$, 70 $g.L^{-1}$, 70 $g.L^{-1}$, 90 $g.L^{-1}$, 100 $g.L^{-1}$, 150 $g.L^{-1}$, 200 $g.L^{-1}$,250 $g.L^{-1}$, 300 $g.L^{-1}$, 350 $g.L^{-1}$, 400 $g.L^{-1}$, 500 $g.L^{-1}$, 600 $g.L^{-1}$, 700 $g.L^{-1}$, 800 $g.L^{-1}$, 900 $g.L^{-1}$, 1000 $g.L^{-1}$, or more.

The thus purified protein may be lyophilised. Alternatively it may be formulated with a carrier or diluent, and optionally presented in a unit form.

It is preferred that the protein is isolated to achieve a pharmaceutically acceptable level of purity. A protein has a pharmaceutically acceptable level of purity if it is essentially pyrogen free and can be administered in a pharmaceutically efficacious amount without causing medical effects not associated with the activity of the protein.

The resulting protein may be used for any of its known utilities, which, in the case of albumin, include i.v. administration to patients to treat severe burns, shock and blood loss, supplementing culture media, and as an excipient in formulations of other proteins.

Although it is possible for a therapeutically useful desired protein obtained by a process of the invention to be administered alone, it is preferable to present it as a pharmaceutical formulation, together with one or more acceptable carriers or diluents. The carrier(s) or diluent(s) must be "acceptable" in the sense of being compatible with the desired protein and not deleterious to the recipients thereof. Typically, the carriers or diluents will be water or saline which will be sterile and pyrogen free.

Optionally the thus formulated protein will be presented in a unit dosage form, such as in the form of a tablet, capsule, injectable solution or the like.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows a plasmid map of the 2μm plasmid.

Figure 2 shows a plasmid map of pSAC35.

Figure 3 shows some exemplified FLP insertion sites.

Figures 4 to 8, 10, 11, 13 to 32 and 36 shows maps of various plasmids.

Figure 9 shows the DNA fragment from pDB2429 containing the *PDI1* gene.

Figure 12 shows some exemplified REP2 insertion sites.

Figure 33 shows table 3 as referred to in the Examples.

Figure 34 shows the sequence of SEQ ID NO: 1.

Figure 35 shows the sequence of SEQ ID NO: 2.

EXAMPLES

These example describes the insertion of additional DNA sequences into a number of positions, defined by restriction endonuclease sites, within the US-region of a 2μm-family plasmid, of the type shown in Figure 2 and generally designated pSAC35, which includes a β-lactamase gene (for ampicillin resistance, which is lost from the plasmid following transformation into yeast), a *LEU2* selectable marker and an oligonucleotide linker, the latter two of which are inserted into a unique *Sna*BI-site within the UL-region of the 2μm-family disintegration vector, pSAC3 (see EP 0 286 424). The sites chosen were towards the 3'-ends of the

REP2 and *FLP* coding regions or in the downstream inverted repeat sequences. Short synthetic DNA linkers were inserted into each site, and the relative stabilities of the modified plasmids were compared during growth on non-selective media. Preferred sites for DNA insertions were identified. Insertion of larger DNA fragments containing "a gene of interest" was demonstrated by inserting a DNA fragment containing the *PDI1* gene into the *Xcm*I-site after *REP2*.

EXAMPLE 1

Insertion of Synthetic DNA Linker into XcmI-Sites in the Small Unique Region of pSAC35

Sites assessed initially for insertion of additional DNA into the US-region of pSAC35, were the *Xcm*I-sites in the 599-bp inverted repeats. One *Xcm*I-site cuts 51-bp after the *REP2* translation termination codon, whereas the other *Xcm*I-site cuts 127-bp before the end of the *FLP* coding sequence, due to overlap with the inverted repeat (see Figure 3).

The sequence inserted was a 52-bp linker made by annealing 0.5mM solutions of oligonucleotides CF86 and CF87. This DNA linker contained a core region "*Sna*BI-*Pac*I-*Fse*I/*Sfi*I-*Sma*I-*Sna*BI", which encoded restriction sites absent from pSAC35.

<u>*Xcm*I Linker (CF86+CF87)</u>

```
                                                 SfiI
                                            --------------
                                  PacI                          SnaBI
                               ---------                       -------
                   SnaBI                    FseI       SmaI
                  -------                 --------    ------
CF86   GGAGTGGTA CGTATTAATT AAGGCCGGCC AGGCCCGGGT ACGTACCAAT TGA
CF87  TCCTCACCAT GCATAATTAA TTCCGGCCGG TCCGGGCCCA TGCATGGTTA AC
```

Plasmid pSAC35 was partially digested with *Xcm*I, the linear 11-kb fragment was isolated from a 0.7%(w/v) agarose gel, ligated with the CF86/CF87 *Xcm*I linker (neat, 10^{-1} and 10^{-2} dilutions) and transformed into *E. coli* DH5α. Ampicillin resistant transformants were selected and screened for the presence of plasmids that could be linearised by *Sma*I digestion. Restriction enzyme analysis identified pDB2688 (Figure 4) with the linker cloned into the *Xcm*I-site after *REP2*. DNA sequencing using oligonucleotides primers CF88, CF98 and CF99 (Table 1) confirmed the insertion contained the correct linker sequence.

Table 1: Oligonucleotide sequencing primers:

Primer	Description	Sequence
CF88	*REP2* primer, 20mer	5'-ATCACGTAATACTTCTAGGG-3'
CF98	*REP2* primer, 20mer	5'-AGAGTGAGTTGGAAGGAAGG-3'
CF99	*REP2* primer, 20mer	5'-AGCTCGTAAGCGTCGTTACC-3'
CF90	*FLP* primer, 20mer	5'-CTAGTTTCTCGGTACTATGC-3'
CF91	*FLP* primer, 20mer	5'-GAGTTGACTAATGTTGTGGG-3'
CF100	*FLP* primer, 20mer	5'-AAAGCTTTGAAGAAAAATGC-3'
CF101	*FLP* primer, 20mer	5'-GCAAGGGGTAGGATCGATCC-3'
CF123	pDB2783 MCS, 24mer	5'-ATTCGAGCTCGGTACCTACGTACT-3'

Primer	Description	Sequence
CF126	pDB2783 MCS, 24mer	5'-CCCGGGCACGTGGGATCCTCTAGA-3'
M13-Forward	pDB2783 MCS, 17mer	5'-GTAAAACGACGGCCAGT-3'
M13-Reverse	pDB2783 MCS, 16mer	5'-AACAGCTATGACCATG-3'

Restriction enzyme analysis also identified pDB2689 (Figure 5), with the linker cloned into the *XcmI*-site in the *FLP* gene. However, the linker in pDB2689 was shown by DNA sequencing using primers CF90 and CF91 to have a missing G:C base-pair within the *FseI/SfiI* site (marked above in bold in the CF86+CF87 linker). This generated a coding sequence for a mutant Flp-protein, with 39 C-terminal amino acid residues replaced by 56 different amino acids before the translation termination codon.

The missing base-pair in the pDB2689 linker sequence was corrected to produce pDB2786 (Figure 6). To achieve this, a 31-bp 5'-phosphorylated *Sna*BI-linker was made from oligonucleotides CF104 and CF105. This was ligated into the *Sna*BI site of pDB2689, which had previously been treated with calf intestinal alkaline phosphatase. DNA sequencing with primers CF90, CF91, CF100 and CF101 confirmed the correct DNA linker sequence in pDB2786. This generated a coding sequence for a mutant Flp-protein, with 39 C-terminal residues replaced by 14 different residues before translation termination.

*Sna*BI Linker (CF104+CF105)

```
                          SfiI
                    ---------------
                        FseI
                    ---------
                PacI                SmaI
            ----------          ------
CF104 Pi-GTATTAATTA AGGCCGGCCA GGCCCGGGTA C
CF105     CATAATTAAT TCCGGCCGGT CCGGGCCCAT G-Pi
```

An additional plasmid, pDB2798 (Figure 7), was also produced by ligation of the *Sna*BI linker in the opposite direction to pDB2786. The linker sequence in pDB2798 was confirmed by DNA sequencing. Plasmid pDB2798 contained a coding sequence for a mutant Flp-protein, with 39 C-terminal residues replaced by 8 different residues before translation termination.

A linker was also cloned into the *Xcm*I-site in the *FLP* gene to truncate the Flp protein at the site of insertion. The linker used was a 45-bp 5'-phosphorylated *Xcm*I-linker made from oligonucleotides CF120 and CF121.

*Xcm*I Linker (CF120+CF121)

```
                                    SfiI
                              ---------------
                        PacI                      SnaBI
                    ---------                   -------
                SnaBI           FseI        SmaI
            --------        ---------    ------
CF120     Pi-GTAATAATA CGTATTAATT AAGGCCGGCC AGGCCCGGGT ACGTAA
CF121       TCATTATTAT GCATAATTAA TTCCGGCCGG TCCGGGCCCA TGCAT-Pi
```

This CF120/CF121 *Xcm*I linker was ligated with 11-kb pSAC35 fragments produced by partial digestion with *Xcm*I, followed by treatment with calf intestinal alkaline phosphatase. Analysis of ampicillin resistant *E. coli* DH5α transformants identified clones containing pDB2823 (Figure 8). DNA sequencing with primers CF90, CF91, CF100 and CF101 confirmed the linker sequence in pDB2823. Translation termination within the linker inserted would result in the production of Flp (1-382), which lacked 41 C-terminal residues.

The impact on plasmid stability from insertion of linker sequences into the *Xcm*I-sites within the US-region of pSAC35 was assessed for pDB2688 and pDB2689. Plasmid stability was determined in a *S. cerevisiae* strain by loss of the *LEU2* marker during non-selective grown on YEPS. The same yeast strain, transformed with pSAC35, which is structurally similar to pSAC3, but contains additional DNA inserted at the *Sna*BI site that contained a *LEU2* selectable marker (Chinery & Hinchliffe, 1989, *Curr. Genet.*, **16**, 21), was used as the control.

The yeast strain was transformed to leucine prototrophy using a modified lithium acetate method (Sigma yeast transformation kit, YEAST-1, protocol 2; (Ito *et al*, 1983, *J. Bacteriol.*, **153**, 163; Elble, 1992, *Biotechniques*, **13**, 18)). Transformants were selected on BMMD-agar plates, and were subsequently patched out on BMMD-agar plates. Cryopreserved trehalose stocks were prepared from 10mL BMMD shake flask cultures (24 hrs, 30^{0}C, 200rpm).

The composition of YEPD and BMMD is described by Sleep *et al.*, 2002, *Yeast* **18**, 403. YEPS and BMMS are similar in composition to YEPD and BMMD accept that 2% (w/v) sucrose was substituted for the 2% (w/v) glucose as the sole initial carbon source.

For the determination of plasmid stability a 1mL cryopreserved stock was thawed and inoculated into 100mL YEPS (initial $OD_{600} \approx 0.04$–0.09) in a 250mL conical flask and grow for approximately 72 hours (70-74 hrs) at 30°C in an orbital shaker (200 rpm, Innova 4300 incubator shaker, New Brunswick Scientific).

Samples were removed from each flask, diluted in YEPS-broth (10^{-2} to 10^{-5} dilution), and 100µL aliquots plated in duplicate onto YEPS-agar plates. Cells were grown at 30°C for 3-4 days to allow single colonies to develop. For each yeast stock analysed, 100 random colonies were patched in replica onto BMMS-agar plates followed by YEPS-agar plates. After growth at 30°C for 3-4 days the

percentage of colonies growing on both BMMS-agar plates and YEPS-agar plates was determined as the measure of plasmid stability.

In the above analysis to measure the loss of the *LEU2* marker from transformants, pSAC35 and pDB2688 appeared to be 100% stable, whereas pDB2689 was 72% stable. Hence, insertion of the linker into the *Xcm*I-site after *REP2* had no apparent effect on plasmid stability, despite altering the transcribed sequence and disrupting the homology between the 599-bp inverted repeats. Insertion of the linker at the *Xcm*I-site in *FLP* also resulted in a surprisingly stable plasmid, despite both disruption of the inverted repeat and mutation of the Flp protein.

EXAMPLE 2

Insertion of the PDI1 Gene into the XcmI Linker of pDB2688

The insertion of a large DNA fragment into the US-region of 2μm-like vectors was demonstrated by cloning the *S. cerevisiae PDI1* gene into the *Xcm*I-linker of pDB2688. The *PDI1* gene (Figure 9) was cloned on a 1.9-kb *Sac*I-*Spe*I fragment from a larger *S. cerevisiae* genomic DNA fragment containing the *PDI1* gene, which had been cloned into YIplac211 (Gietz & Sugino, 1988, *Gene*, 74, 527-534), which had been treated with T4 DNA polymerase to fill the *Spe*I 5'-overhang and remove the *Sac*I 3'-overhang. This *PDI1* fragment included 212-bp of the *PDI1* promoter upstream of the translation initiation codon, and 148-bp downstream of the translation termination codon. This was ligated with *Sma*I linearised/calf intestinal alkaline phosphatase treated pDB2688, to create plasmid pDB2690 (Figure 10), with the *PDI1* gene transcribed in the same direction as *REP2*. A *S. cerevisiae* strain was transformed to leucine prototrophy with pDB2690.

An expression cassette for a human transferrin mutant (N413Q, N611Q) was subsequently cloned into the *Not*I-site of pDB2690 to create pDB2711 (Figure 11). The expression cassette in pDB2711 contains the *S. cerevisiae PRB1* promoter, an HSA/MFα fusion leader sequence (EP 387319; Sleep *et al*, 1990, *Biotechnology (N.Y.)*, **8**, 42) followed by a coding sequence for the human transferrin mutant (N413Q, N611Q) and the *S. cerevisiae ADH1* terminator. Plasmid pDB2536 (Figure 36) was constructed similarly by insertion of the same expression cassette into the *Not*I-site of pSAC35.

The advantage of inserting "genes of interest" into the US-region of 2μm-vectors was demonstrated by the approximate 7-fold increase in recombinant transferrin N413Q, N611Q secretion during fermentation of yeast transformed with pDB2711, compared to the same yeast transformed with pDB2536. An approximate 15-fold increase in recombinant transferrin N413Q, N611Q secretion was observed in shake flask culture (data not shown).

The relative stabilities of plasmids pDB2688, pDB2690, pDB2711, pDB2536 and pSAC35 were determined in the same yeast strain grown in YEPS media, using the method described above (Table 2).

Table 2: Summary of plasmid stability data for *PDI1* insertion into the small unique region of pSAC35. Data from 3 days growth in non-selective shake flask culture before plating on YEPS-agar.

Plasmid	Insertion Site(s)	Additional Details	Relative Stability
pSAC35	-	-	100%
pDB2688	*Xcm*I	Linker in Inverted Repeat	100%
pDB2690	*Xcm*I	*PDI1* in *Xcm*I Linker	32%
pDB2711	*Xcm*I, *Not*I	*PDI1* in *Xcm*I Linker rTf Cassette at *Not*I	10%
pDB2536	*Not*I	rTf Cassette at *Not*I	17%

In this analysis, pDB2690 was 32% stable, compared to 100% stability for pDB2688 without the *PDI1* insert. This decrease in plasmid stability was less than the decrease in plasmid stability observed with pDB2536, due to insertion of the rTF (N413Q, N611Q) expression cassette into the *Not*I-site within the large unique region of pSAC35 (Table 2).

Furthermore, selective growth in minimal media during high cell density fermentations could overcome the increased plasmid instability due to the *PDI1* insertion observed in YEPS medium, as the rTF (N413Q, N611Q) yield from the same yeast transformed with pDB2711 did not decrease compared to that achieved from the same yeast transformed with pDB2536.

EXAMPLE 3

Insertion of DNA Linkers into the REP2 Gene and Downstream Sequences in the Inverted Repeat of pSAC35

To define the useful limits for insertion of additional DNA into the *REP2* gene and sequences in the inverted repeat downstream of it, further linkers were inserted into pSAC35. Figure 12 indicates the restriction sites used for these insertions and the effects on the Rep2 protein of translation termination at these sites.

The linker inserted at the *Xmn*I-site in *REP2* was a 44-bp sequence made from oligonucleotides CF108 and CF109.

<u>*Xmn*I Linker (CF108+CF109)</u>

```
                                         SfiI
                                   --------------
                         PacI                                SnaBI
                     ----------                             ---|---
              SnaBI                 FseI         SmaI          |
             -------              --------      ------         |
CF108   ATAATAATAC GTATTAATTA AGGCCGGCCA GGCCCGGGTA CGTA
CF109   TATTATTATG CATAATTAAT TCCGGCCGGT CCGGGCCCAT GCAT
```

To avoid insertion into other *Xmn*I-sites in pSAC35, the 3,076-bp *Xba*I fragment from pSAC35 that contained the *REP2* and *FLP* genes was first sub-cloned into the *E. coli* cloning vector pDB2685 (Figure 13) to produce pDB2783 (Figure 14).

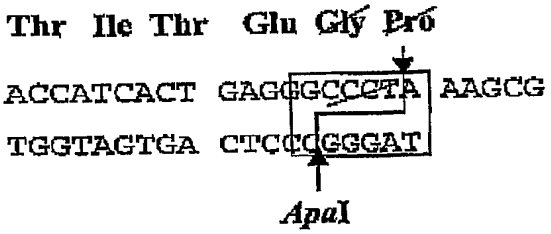
Plasmid pDB2685 is a pUC18-like cloning vector derived from pCF17 containing apramycin resistance gene aac(3)IV from *Klebsiella pneumoniae* (Rao *et al*, 1983, *Antimicrob. Agents Chemother.*, **24**, 689) and multiple cloning site from pMCS5 (Hoheisel, 1994, *Biotechniques*, **17**, 456). pCF17 was made from pIJ8600 (Sun *et al.*, 1999, *Microbiology*, **145**(9), 2221-7) by digestion with *Eco*RI, *Nhe*I and the Klenow fragment of DNA polymerase I, and self-ligation, followed by isolation from the reaction products by transformation of competent *E. coli* DH5α cells and

selection with apramycin sulphate. Plasmid pDB2685 was constructed by cloning a 439bp *Ssp*I-*Swa*I fragment from pMCS5 into pCF17, which had been cut with *Msc*I and treated with calf intestinal alkaline phosphatase. Blue/white selection is not dependant on IPTG induction.

Plasmid pDB2783 was linearised with *Xmn*I and ligated with the CF108/CF109 *Xmn*I-linker to produce pDB2799 (Figure 15) and pDB2780 (not shown). Plasmid pDB2799 contained the CF108/CF109 *Xmn*I linker in the correct orientation for translation termination at the insertion site to produce Rep2 (1-244), whereas pDB2780 contained the linker cloned in the opposite orientation. DNA sequencing with primers CF98 and CF99 confirmed the correct linker sequences.

The 3,120bp *Xba*I fragment from pDB2799 was subsequently ligated with a 7,961-bp pSAC35 fragment which had been produced by partial *Xba*I digestion and treatment with calf intestinal alkaline phosphatase, to create plasmid pDB2817 (B-form) and pDB2818 (A-form) disintegration vectors (Figures 16 and 17 respectively).

Insertion of linkers at the *Apa*I-site in pSAC35 was performed with and without 3'-5' exonuclease digestion by T4 DNA polymerase. This produced coding sequences for either Rep2 (1-271) or Rep2 (1-269) before translation termination. In the following figure, the sequence GGCC marked with diagonal lines was deleted from the 3'-overhang produced after ApaI digestion resulting in removal of nucleotides from the codons for Glycine-170 (GGC) and Proline-171.



The linker inserted at the *Apa*I-site without exonuclease digestion was a 50-bp 5'-phosphorylated linker made from oligonucleotides CF116 and CF117.

<u>*Apa*I-Linker (CF116+CF117)</u>

```
                                            SfiI
                                       -------------
                              PacI                          SnaBI
                           ---------                        ------
                  SnaBI                 FseI        SmaI
                  ------              ---------    -------
CF116     Pi-CTTAAT AATACGTATT AATTAAGGCC GGCCAGGCCC GGGTACGTAG GGCC
CF117    CCGGGAATTA TTATGCATAA TTAATTCCGG CCGGTCCGGG CCCATGCATC-Pi
```

This was ligated with pSAC35, which had been linearised with *Apa*I and treated with calf intestinal alkaline phosphatase, to produce pDB2788 (Figure 18) and pDB2789 (not shown). Within pDB2788, the linker was in the correct orientation for translation termination after proline-271, whereas in pDB2789 the linker was in the opposite orientation.

The linker inserted at the *Apa*I-site with exonuclease digestion by T4 DNA polymerase was a 43-bp 5'-phosphorylated linker made from oligonucleotides CF106 and CF107, which was called the core termination linker.

<u>Core Termination-Linker (CF106+CF107)</u>

```
                                              SfiI
                                        --------------
                            PacI                            SnaBI
                          --------                         -------
                SnaBI                  FseI       SmaI
               -------               --------    ------
CF106     Pi-TAATAATACG TATTAATTAA GGCCGGCCAG GCCCGGGTAC GTA
CF107        ATTATTATGC ATAATTAATT CCGGCCGGTC CGGGCCCATG CAT-Pi
```

The core termination linker was ligated with pSAC35, which had been linearised with *Apa*I, digested with T4 DNA polymerase and treated with calf intestinal alkaline phosphatase. This ligation produced pDB2787 (Figure 19) with the linker cloned in the correct orientation for translation termination after glutamate-269.

The correct DNA sequences were confirmed in all clones containing the *Apa*I-linkers, using oligonucleotide primers CF98 and CF99.

The core termination linker (CF106+CF107) was also used for insertion into the *Fsp*I-sites of pDB2783 (Figure 14). The core termination linker (CF106+CF107) was ligated into pDB2783 linearised by partial *Fsp*I digestion, which had been treated with calf intestinal alkaline phosphatase. Plasmids isolated from apramycin resistant *E. coli* DH5α transformants were screened by digestion with *Fsp*I, and selected clones were sequenced with M13 forward and reverse primers.

Plasmid pDB2801 (not shown) was identified containing two copies of the linker cloned in the correct orientation (with the *Pac*I-site nearest the *REP2* gene). The extra copy of the linker was subsequently removed by first deleting a 116-bp *Nru*I-*Hpa*I fragment containing an *Fse*I-site from the multiple cloning site region, followed by digestion with *Fse*I and re-ligation to produce pDB2802 (Figure 20). DNA sequencing using oligonucleotide CF126 confirmed the correct linker sequence.

The 3,119-bp pDB2802 *Xba*I fragment was subsequently ligated with a 7,961-bp pSAC35 fragment produced by partial *Xba*I digestion and treatment with calf intestinal alkaline phosphatase to create pDB2805 (B-form) and pDB2806 (A-form) disintegration vectors (Figures 21 and 22, respectively).

EXAMPLE 4

Insertion of DNA Linkers into the FLP Gene and Downstream Sequences in the Inverted Repeat of pSAC35

DNA linkers were inserted into pSAC35 to define the useful limits for insertion of additional DNA into the *FLP* gene and sequences downstream in the inverted

repeat. Figure 3 indicates the restriction sites used for these insertions and the affects on the Flp protein of translation termination at these sites.

The linker inserted at the *BcI*I-site was a 49-bp 5'-phosphorylated linker made from oligonucleotides CF118 and CF119.

<u>*BcI*I Linker (CF118+CF119)</u>

```
                                   SfiI
                           -------------
                       PacI                  SnaBI
                     --------               ------
                SnaBI           FseI    SmaI
               ------      --------    ------
CF118 Pi-GATCACTAATAATACGTATTAATTAAGGCCGGCCAGGCCCGGGTACGTA
CF119        TGATTATTATGCATAATTAATTCCGGCCGGTCCGGGCCCATGCATCTAG-Pi
```

Due to Dam-methylation of the *BcI*I-site in pSAC35, the *BcI*I-linker was cloned into non-methylated pSAC35 DNA, which had been isolated from the *E. coli* strain ET12567 pUZ8002 (MacNeil *et al*, 1992, *Gene*, **111**, 61; Kieser *et al*, 2000, *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich). Plasmid pSAC35 was linearised with *BcI*I, treated with calf intestinal alkaline phosphatase, and ligated with the *BcI*I-linker to create pDB2816 (Figure 23). DNA sequencing with oligonucleotide primers CF91 and CF100 showed that three copies of the *BcI*I-linker were present in pDB2816, which were all in the correct orientation for translational termination of Flp after histidine-353.

Digestion of pDB2816 with *Pac*I followed by self-ligation, was performed to produce pDB2814 and pDB2815, containing one and two copies of the *BcI*I-linker respectively (Figures 24 and 25). The DNA sequences of the linkers were confirmed using primers CF91 and CF100. In *S. cerevisiae* a truncated Flp (1-353) protein will be produced by yeast transformed with pDB2814, pDB2815 or pDB2816.

An additional plasmid pDB2846 (data not shown) was also produced by ligation of a single copy of the *BclI*-linker in the opposite orientation to pDB2814. This has the coding sequence for the first 352-residues from Flp followed by 14 different residues before translation termination.

The linker inserted at the *HgaI*-site was a 47-bp 5'-phosphorylated linker made from oligonucleotides CF114 and CF115.

HgaI Linker (CF114+CF115)

```
                                        SfiI
                                    ------------
                              PacI                    SnaBI
                            --------                  -----
                        SnaBI          FseI       SmaI
                      -------        --------   ------
CF114   Pi-AGTACTATAATACGTATTAATTAAGGCCGGCCAGGCCCGGGTACGTA
CF115           ATATTATGCATAATTAATTCCGGCCGGTCCGGGCCCATGCATTCATG-Pi
```

The *HgaI*-linker was ligated with pDB2783, which had been linearised by partial *HgaI* digestion and treated with calf intestinal alkaline phosphatase to create pDB2811 (Figure 26). DNA sequencing with oligonucleotides CF90, CF91 and CF100 confirmed the correct linker insertion.

The 3,123-bp *XbaI* fragment from pDB2811 was subsequently ligated with the 7,961-bp pSAC35 fragment, produced by partial *XbaI* digestion and treatment with calf intestinal alkaline phosphatase to produce pDB2812 (B-form) and pDB2813 (A-form) disintegration vectors containing DNA inserted at the *HgaI*-site (Figures 27 and 28, respectively).

Plasmids pDB2803 and pDB2804 (Figures 29 and 30, respectively) with the core termination linker (CF106+CF107) inserted at the *FspI* after *FLP*, were isolated by the same method used to construct pDB2801. The correct linker insertions were confirmed by DNA sequencing. Plasmid pDB2804 contained the linker inserted

in the correct orientation (with the *Pac*I-site closest to the *FLP* gene), whereas pDB2803 contained the linker in the opposite orientation.

The pDB2804 3,119-bp *Xba*I fragment was ligated with the 7,961-bp pSAC35 fragment produced by partial *Xba*I digestion and treatment with calf intestinal alkaline phosphatase to create pDB2807 (B-form) and pDB2808 (A-form) disintegration vectors containing DNA inserted at the *Fsp*I-site after *FLP* (Figures 31 and 32 respectively).

EXAMPLE 5

Relative Stabilities of the LEU2 Marker in Yeast Transformed with pSAC35-Like Plasmids Containing DNA Linkers Inserted into the Small Unique Region and Inverted Repeats

A *S. cerevisiae* strain was transformed with the pSAC35-like plasmids containing DNA linkers inserted into the US-region and inverted repeats. Cryopreserved trehalose stocks were prepared for testing plasmid stabilities (Table 3). Plasmid stabilities were analysed as described above for linkers inserted at the *Xcm*I-sites in pSAC35. Duplicate flasks were set up for each insertion site analysed. In addition, to the analysis of colonies derived from cells after 3-days in shake flake culture, colonies were grown and analysed from cells with a further 4-days shake flask culture. For this, 100µL samples were removed from each 3-day old flask and sub-cultured in 100mL YEPS broth for a further period of approximately 96 hours (94-98 hrs) at 30.0°C in an orbital shaker, after which single colonies were obtained and analysed for loss of the *LEU2* marker. In this case analysis was restricted to a single flask from selected strains, for which 50 colonies were picked. The overall results are summarised in Table 4.

Table 4: Summary of plasmid stability data for DNA insertions into pSAC35

Set 1 represents data from 3 days in non-selective shake flask culture.

Set 2 represents data from 7 days in non-selective shake flask culture.

A) *REP2* Insertion Sites

Plasmid(s)	Insertion Site	Additional Details	Relative Stability	
			Set 1	Set 2
pSAC35	-	Control	99%	100%
pDB2817 & pDB2818	*Xmn*I	REP2 (1-244)	39%	16%
pDB2787	*Apa*I/T4 pol.	REP2 (1-269)	45%	0%
pDB2788	*Apa*I	REP2 (1-271)	33%	0%
pDB2688	*Xcm*I	Inverted Repeat	100%	100%
pDB2805 & pDB2806	*Fsp*I	Inverted Repeat	100%	100%

B) *FLP* Insertion Sites

Plasmid(s)	Insertion Site	Additional Details	Relative Stability	
			Set 1	Set 2
pDB2814	*Bcl*I	FLP (1-353)	67%	64%
pDB2823	*Xcm*I	FLP (1-382)	64%	53%
pDB2812 & pDB2813	*Hga*I	Inverted Repeat	100%	100%
pDB2808	*Fsp*I	Inverted Repeat	100%	100%

All of the modified pSAC35 plasmids were able to transform yeast to leucine prototrophy, indicating that despite the additional DNA inserted within the functionally crowded regions of 2µm DNA, all could replicate and partition in *S. cerevisiae*. This applied to plasmids with 43-52 base-pair linkers inserted at all the sites in the 2µm US-region, as well as the larger DNA insertion containing the *PDI1* gene.

For the linker insertion sites, data was reproducible between both experiments and duplicates. All sites outside *REP2* or *FLP* open reading frames, but within inverted repeats appeared to be 100% stable under the test conditions used. Plasmid instability (i.e. plasmid loss) was observed for linkers inserted into sites within the *REP2* or *FLP* open reading frames. The observed plasmid instability of *REP2* insertions was greater than for *FLP* insertions. For the *REP2* insertions, loss of the *LEU2* marker continued with the extended growth period in non-selective media, whereas there was little difference for the *FLP* insertions.

Insertions into the *REP2* gene produced Rep2 polypeptides truncated within a region known to function in self-association and binding to the STB-locus of 2μm (Sengupta *et al*, 2001, *J. Bacteriol.*, **183**, 2306).

Insertions into the *FLP* gene resulted in truncated Flp proteins. All the insertion sites were after tyrosine-343 in the C-terminal domain, which is essential for correct functioning of the Flp protein (Prasad *et al*, 1987, *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 2189; Chen *et al*, 1992, *Cell*, **69**, 647; Grainge *et al*, 2001, *J. Mol. Biol.*, **314**, 717).

None of the insertions into the inverted repeat regions resulted in plasmid instability being detected, except for the insertion into the *FLP Xcm*I-site, which also truncated the Flp protein product. The insertions at the *FspI*-sites in the inverted repeat regions were the closest to the FRT (Flp recognition target) regions, important for plasmid replication.

pSAC35-like plasmids have been constructed with 43-52 base-pair DNA linkers inserted into the *REP2* open reading frame, or the *FLP* open reading frame or the inverted repeat sequences. In addition, a 1.9-kb DNA fragment containing the *PDI1* gene was inserted into a DNA linker at the *Xcm*I-site after *REP2*.

All of the pSAC35-like vectors with additional DNA inserted were able to transform yeast to leucine prototrophy. Therefore, despite inserting DNA into functionally crowded regions of 2μm plasmid DNA, the plasmid replication and partitioning mechanisms had not been abolished.

Determination of plasmid stability by measuring loss of the *LEU2* selectable marker during growth in non-selective medium indicated that inserting DNA linkers into the inverted repeats had not destabilised the plasmid, whereas plasmid stability had been reduced by insertions into the *REP2* and *FLP* open reading frames. However, despite a reduction in plasmid stability under non-selective media growth conditions when insertions were made into the *REP2* and *FLP* open reading frames at some positions defined by the first and second aspects of the invention, the resulting plasmid nevertheless has a sufficiently high stability for use in yeast when grown on selective media.

CLAIMS

1. A 2μm-family plasmid comprising a polynucleotide sequence insertion, deletion and/or substitution between the first base after the last functional codon of at least one of either a *REP2* gene or an *FLP* gene and the last base before the FRT site in an inverted repeat adjacent to said gene.

2. The 2μm-family plasmid of Claim 1 wherein, other than the polynucleotide sequence insertion, deletion and/or substitution, the *FLP* gene and/or the *REP2* gene has the sequence of a *FLP* gene and/or a *REP2* gene, respectively, derived from a naturally occurring 2μm-family plasmid.

3. The 2μm-family plasmid of Claim 1 wherein the naturally occurring 2μm-family plasmid is selected from pSR1, pSB3 or pSB4 as obtained from *Zygosaccharomyces rouxii*, pSB1 or pSB2 both as obtained from *Zygosaccharomyces bailli*, pSM1 as obtained from *Zygosaccharomyces fermentati*, pKD1 as obtained from *Kluyveromyces drosophilarum*, pPM1 as obtained from *Pichia membranaefaciens*, and the 2μm plasmid as obtained from *Saccharomyces cerevisiae*.

4. The 2μm-family plasmid of Claim 2 or 3 wherein the sequence of the inverted repeat adjacent to said *FLP* and/or *REP2* gene is derived from the sequence of the corresponding inverted repeat in the same naturally occurring 2μm-family plasmid as the sequence from which the gene is derived.

5. The 2μm-family plasmid of any one of Claims 2 to 4 wherein the naturally occurring 2μm-family plasmid is the 2μm plasmid as obtained from *Saccharomyces cerevisiae*.

6. The 2μm-family plasmid of Claim 5 wherein the polynucleotide sequence insertion, deletion and/or substitution occurs at a position between the first base of codon 59 of the *REP2* gene and the last base before the FRT site in the adjacent inverted repeat.

7. The 2μm-family plasmid of Claim 5 or 6 wherein, other than the polynucleotide sequence insertion, deletion and/or substitution, the sequence of the *REP2* gene and the adjacent inverted repeat is as defined by SEQ ID NO:1 or variant thereof.

8. The 2μm-family plasmid of any one of Claims 1 to 7 wherein polynucleotide sequence insertion, deletion and/or substitution occurs at a position between the first base of the inverted repeat and the last base before the FRT site.

9. The 2μm-family plasmid of any one of Claims 1 to 7 wherein the polynucleotide sequence insertion, deletion and/or substitution occurs between the first base after the end of the *REP2* coding sequence and the last base before the FRT site, such as at the first base after the end of the *REP2* coding sequence.

10. The 2μm-family plasmid of any one of Claims 1 to 7 wherein, other than the polynucleotide sequence insertion, deletion and/or substitution, the inverted repeat that follows the *REP2* coding sequence has a sequence derived from the corresponding region of the 2μm plasmid as obtained from *Saccharomyces cerevisiae* and preferably the polynucleotide sequence insertion, deletion and/or substitution occurs at an *XcmI* site or an *FspI* site within the inverted repeat.

11. The 2μm-family plasmid of Claim 5 wherein the polynucleotide sequence insertion, deletion and/or substitution occurs at a position between the first base of codon 344 of the *FLP* gene and the last base before the FRT site in the adjacent inverted repeat.

12. The 2μm-family plasmid of Claim 5 or 11 wherein, other than the polynucleotide sequence insertion, deletion and/or substitution, the sequence of the *FLP* coding sequence and the adjacent inverted repeat is as defined by SEQ ID NO:2 or variant thereof.

13. The 2μm-family plasmid of Claim 11 or 12 wherein the polynucleotide sequence insertion, deletion and/or substitution occurs at a position between the first base of the inverted repeat and the last base before the FRT site.

14. The 2μm-family plasmid of Claim 13 wherein the polynucleotide sequence insertion, deletion and/or substitution occurs at a position between the first base after the end of the *FLP* coding sequence and the last base before the FRT site.

15. The 2μm-family plasmid of Claim 14 wherein the polynucleotide sequence insertion, deletion and/or substitution occurs at the first base after the end of the *FLP* coding sequence.

16. The 2μm-family plasmid of any one of Claims 11 to 15 wherein, other than the polynucleotide sequence insertion, deletion and/or substitution, the inverted repeat that follows the *FLP* gene has a sequence derived from the corresponding region of the 2μm plasmid as obtained from *Saccharomyces cerevisiae*, and preferably the polynucleotide sequence insertion, deletion and/or substitution occurs at an *HgaI* site or an *FspI* site within the inverted repeat.

17. The 2μm-family plasmid of any one of the preceding claims comprising polynucleotide sequence insertions, deletions and/or substitutions between the first bases after the last functional codons of both of the *REP2* gene and the *FLP* gene and the last bases before the FRT sites in the inverted repeats adjacent to each of said genes, which polynucleotide sequence insertions, deletions and/or substitutions can be the same or different.

18. The 2μm-family plasmid of any preceding claim additionally comprising a polynucleotide sequence insertion, deletion and/or substitution which is not at a position as defined in any one of the preceding claims.

19. The 2μm-family plasmid of Claim 18 wherein the polynucleotide sequence insertion, deletion and/or substitution occurs within an untranscribed region around an ARS sequence.

20. The 2μm-family plasmid of any one of the preceding claims wherein the, or at least one, polynucleotide sequence insertion, deletion and/or substitution is a polynucleotide sequence insertion.

21. The 2μm-family plasmid of Claim 20 in which the polynucleotide sequence insertion encodes an open reading frame.

22. The 2μm-family plasmid of Claim 21 in which the open reading frame encodes a heterologous protein.

23. The 2μm-family plasmid of Claim 22 in which the heterologous protein comprises the sequence of a protein involved in protein folding, or which has chaperone activity or is involved in the unfolded protein response, albumin, a monoclonal antibody, an etoposide, a serum protein (such as a

blood clotting factor), antistasin, a tick anticoagulant peptide, transferrin, lactoferrin, endostatin, angiostatin, collagens, immunoglobulins, Fab' fragments, F(ab")2, scAb, scFv, interferons, interleukins, IL10, IL11, IL2, interferon α species and sub-species, interferon β species and sub-species, interferon γ species and sub-species, leptin, CNTF, $CNTF_{Ax15}$, IL1-receptor antagonist, erythropoietin (EPO) and EPO mimics, thrombopoietin (TPO) and TPO mimics, prosaptide, cyanovirin-N, 5-helix, T20 peptide, T1249 peptide, HIV gp41, HIV gp120, urokinase, prourokinase, tPA, hirudin, platelet derived growth factor, parathyroid hormone, proinsulin, insulin, glucagon, glucagon-like peptides, insulin-like growth factor, calcitonin, growth hormone, transforming growth factor β, tumour necrosis factor, G-CSF, GM-CSF, M-CSF, FGF, coagulation factors in both pre and active forms, including but not limited to plasminogen, fibrinogen, thrombin, pre-thrombin, pro-thrombin, von Willebrand's factor, α_1-antitrypsin, plasminogen activators, Factor VII, Factor VIII, Factor IX, Factor X and Factor XIII, nerve growth factor, LACI, platelet-derived endothelial cell growth factor (PD-ECGF), glucose oxidase, serum cholinesterase, aprotinin, amyloid precursor protein, inter-alpha trypsin inhibitor, antithrombin III, apo-lipoprotein species, Protein C, Protein S, or a variant or fragment of any of the above.

24. The 2μm-family plasmid of Claim 23 in which the heterologous protein comprises the sequence of albumin, a variant or fragment thereof, or a fusion protein comprising the sequence of any of these.

25. The 2μm-family plasmid of Claim 23 in which the heterologous protein comprises the sequence of transferrin, a variant or fragment thereof, or a fusion protein comprising the sequence of any of these.

26. The 2μm-family plasmid of Claim 23 in which the heterologous protein comprises the sequence of lactoferrin, a variant or fragment thereof, or a fusion protein comprising the sequence of any of these.

27. The 2μm-family plasmid of Claim 23 in which the heterologous protein comprises the sequence of Fc, a variant or fragment thereof, or a fusion protein comprising the sequence of any of these.

28. The 2μm-family plasmid of Claim 23 in which the heterologous protein comprises the sequence of a protein involved in protein folding, or which has chaperone activity or is involved in the unfolded protein response taken from the list *AHA1*, *CCT2*, *CCT3*, *CCT4*, *CCT5*, *CCT6*, *CCT7*, *CCT8*, *CNS1*, *CPR3*, *CPR6*, *EPS1*, *ERO1*, *EUG1*, *FMO1*, *HCH1*, *HSP10*, *HSP12*, *HSP104*, *HSP26*, *HSP30*, *HSP42*, *HSP60*, *HSP78*, *HSP82*, *JEM1*, *MDJ1*, *MDJ2*, *MPD1*, *MPD2*, *PDI1*, *PFD1*, *ABC1*, *APJ1*, *ATP11*, *ATP12*, *BTT1*, *CDC37*, *CPR7*, *HSC82*, *KAR2*, *LHS1*, *MGE1*, *MRS11*, *NOB1*, *ECM10*, *SSA1*, *SSA2*, *SSA3*, *SSA4*, *SSC1*, *SSE2*, *SIL1*, *SLS1*, *UBI4*, *ORM1*, *ORM2*, *PER1*, *PTC2*, *PSE1* and *HAC1* or a truncated intronless *HAC1*.

29. The 2μm-family plasmid of Claim 23 or 28 in which the chaperone is protein disulphide isomerase (PDI).

30. The 2μm-family plasmid of any one of Claims 22 to 29 in which the heterologous protein comprises a secretion leader sequence.

31. The 2μm-family plasmid of Claim 22 in which the heterologous protein comprises the sequence of a bacterial selectable marker and/or a yeast selectable marker.

32. The 2μm-family plasmid of Claim 31 in which the bacterial selectable marker is a β-lactamase gene and/or the yeast selectable marker is a *LEU2* selectable marker.

33. The 2μm-family plasmid according to any one of the preceding claims which plasmid comprises (i) a heterologous sequence encoding a protein; (ii) a heterologous sequence encoding a protein comprising the sequence of a protein involved in protein folding, a chaperone or a protein involved in the unfolded protein response, preferably protein disulphide isomerase; and (iii) a heterologous sequence encoding a protein comprising the sequence of a selectable marker; wherein at least one of the heterologous sequences occurs at a position as defined by any one of Claims 1 to 16.

34. A method of preparing a plasmid as defined by any one of the preceding claims comprising –

(a) providing a plasmid comprising the sequence of a *REP2* gene and the inverted repeat that follows the *REP2* gene, or a *FLP* gene and the inverted repeat that follows the *FLP* gene, in each case the inverted repeat comprising an FRT site;

(b) providing a polynucleotide sequence and inserting the polynucleotide sequence into the plasmid at a position as defined in any one of Claims 1 to 16; and/or

(c) deleting some or all of the nucleotide bases at the positions defined in any one of Claims 1 to 16; and/or

(d) substituting some or all of the nucleotide bases at the positions defined in any one of Claims 1 to 16 with alternative nucleotide bases.

35. A plasmid obtainable by the method of Claim 34.

36. A host cell comprising a plasmid as defined by any one of Claims 1 to 33 and 35.

37. A host cell according to Claim 36 which is a yeast cell.

38. A host cell according to Claim 36 or 37 in which the plasmid is stable as a multicopy plasmid.

39. A host cell according to Claim 38 in which the plasmid is based on pSR1, pSB3 or pSB4 and the yeast cell is *Zygosaccharomyces rouxii*, the plasmid is based on pSB1 or pSB2 and the yeast cell is *Zygosaccharomyces bailli*, the plasmid is based on pSM1 and the yeast cell is *Zygosaccharomyces fermentati*, the plasmid is based on pKD1 and the yeast cell is *Kluyveromyces drosophilarum*, the plasmid is based on pPM1 and the yeast cell is *Pichia membranaefaciens* or the plasmid is based on the 2μm plasmid and the yeast cell is *Saccharomyces cerevisiae* or *Saccharomyces carlsbergensis*.

40. A host cell according to Claim 38 or 39 in which, if the plasmid contains, or is modified to contain, a selectable marker then stability, as measured by the loss of the marker, is at least 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9% or substantially 100% after 5 generations.

41. A method of producing a protein comprising the steps of–

(a) providing a plasmid as defined by any one of Claims 1 to 33 or 35

(b) providing a suitable host cell;

(c) transforming the host cell with the plasmid; and

(d) culturing the transformed host cell in a culture medium;

(e) thereby to produce the protein.

42. A method of producing a protein comprising the steps of providing a host cell as defined by any one of Claims 36 to 40 which host cell comprises a plasmid as defined by any one of Claims 1 to 33 or 35 and culturing the host cell in a culture medium thereby to produce the protein.

43. A method according to Claim 41 or 42 further comprising the step of isolating the thus produced protein from the cultured host cell or the culture medium.

44. A method according to Claim 43 further comprising the step of purifying the thus isolated protein to a commercially acceptable level of purity.

45. A method according to Claim 44 further comprising the step of formulating the thus purified protein with a carrier or diluent, and optionally presenting the thus formulated protein in a unit form.

46. A method according to Claim 43 further comprising the step of purifying the thus isolated protein to a pharmaceutically acceptable level of purity.

47. A method according to Claim 44 further comprising the step of formulating the thus purified protein with a pharmaceutically acceptable carrier or

diluent and optionally presenting the thus formulated protein in a unit dosage form.

5

ABSTRACT

MODIFIED PLASMID AND USE THEREOF

The present invention provides a 2μm-family plasmid comprising a polynucleotide sequence insertion, deletion and/or substitution between the first base after the last functional codon of at least one of either a *REP2* gene or an *FLP* gene and the last base before the FRT site in an inverted repeat adjacent to said gene.

FIGURE 1.

Figure 1

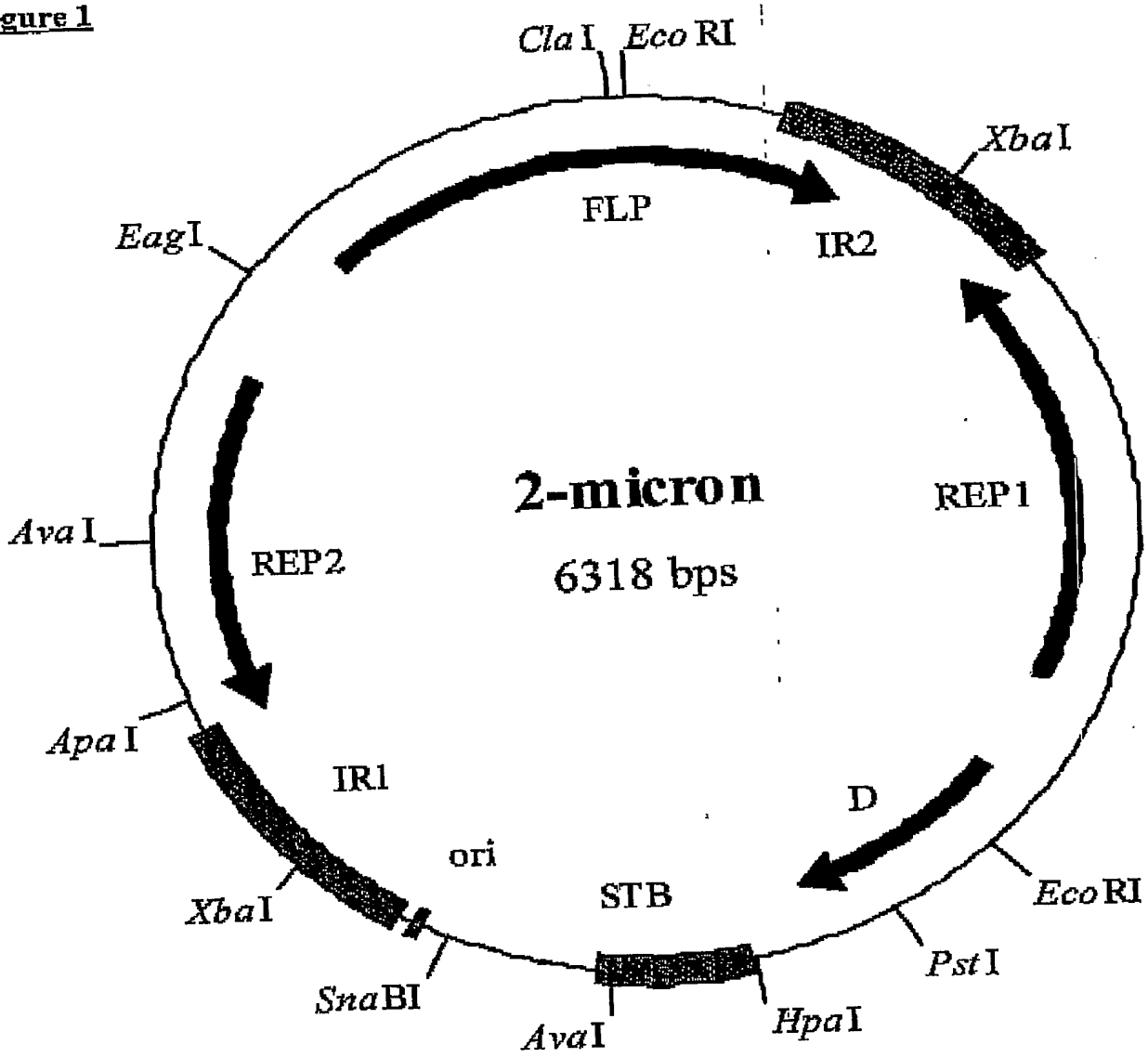


Figure 2

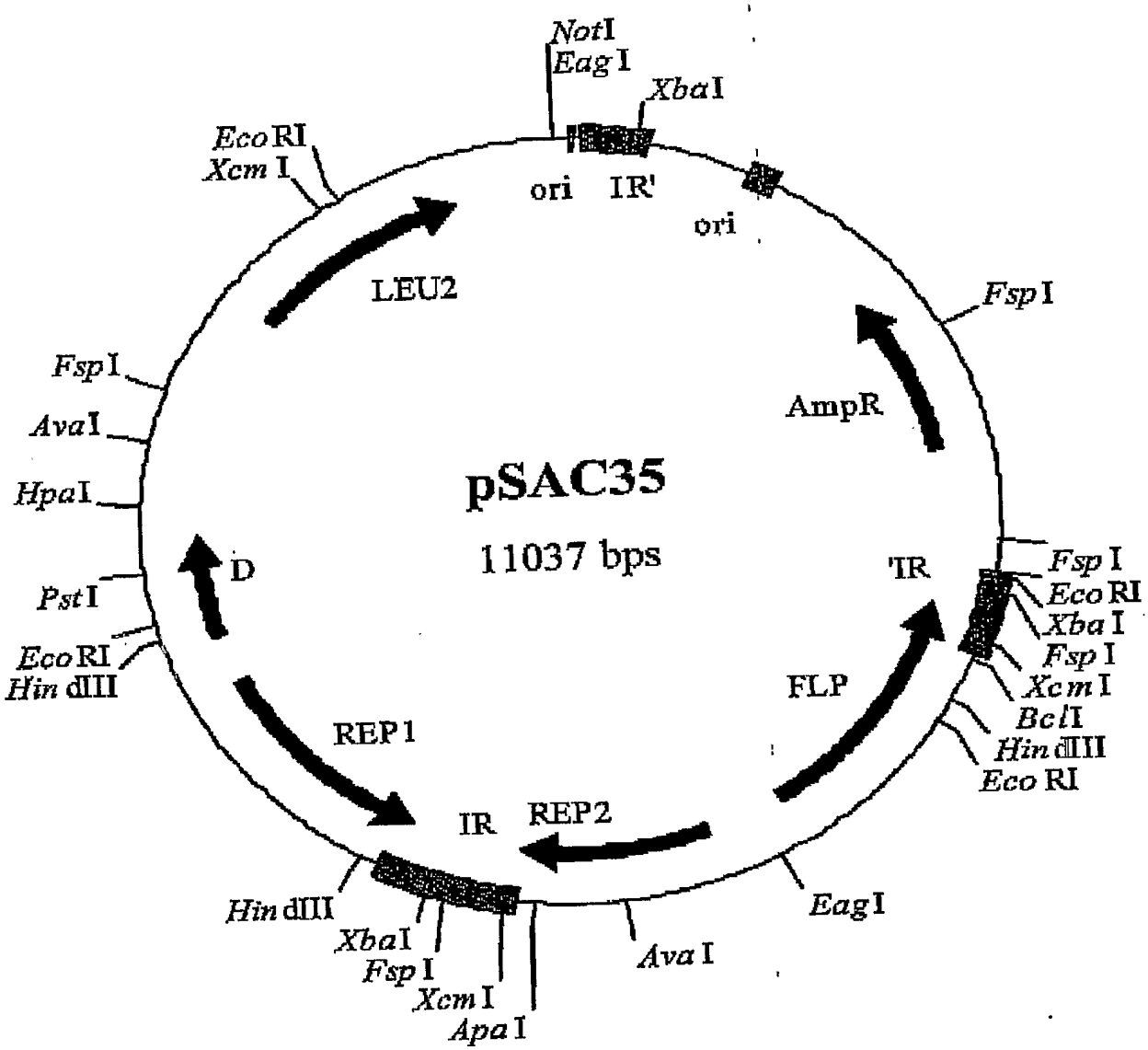
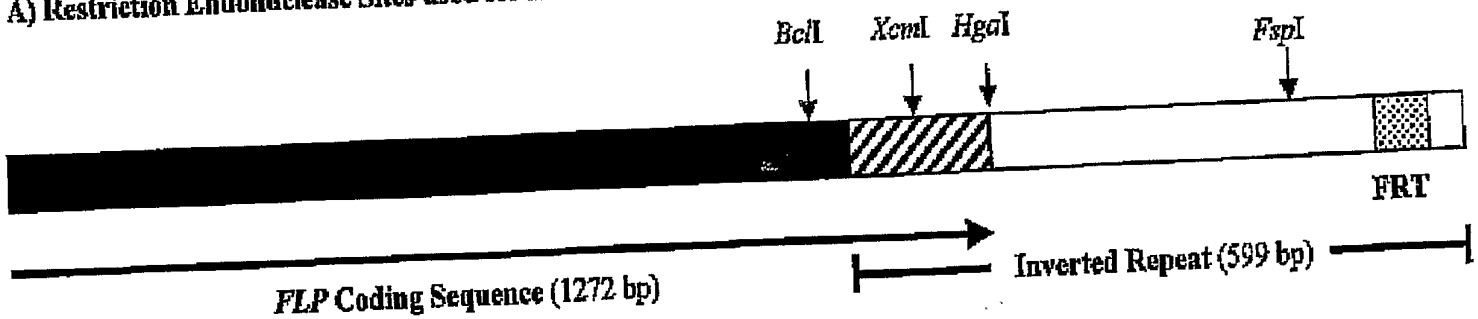


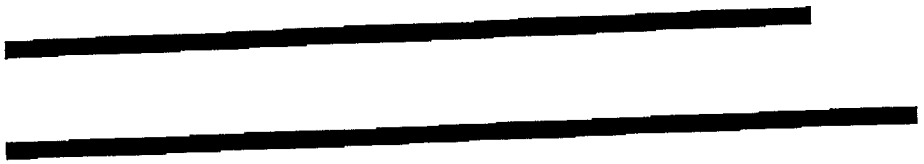
Figure 3

A) Restriction Endonuclease Sites used for DNA Insertions in *FLP* and the *FLP* Inverted Repeat



B) Flp Protein (423 amino acid residues)

C) Truncated FLP Protein Products



Flp (1-353) from termination at *Bcl*I-site

Flp (1-382) from termination at *Xcm*I-site

Figure 4

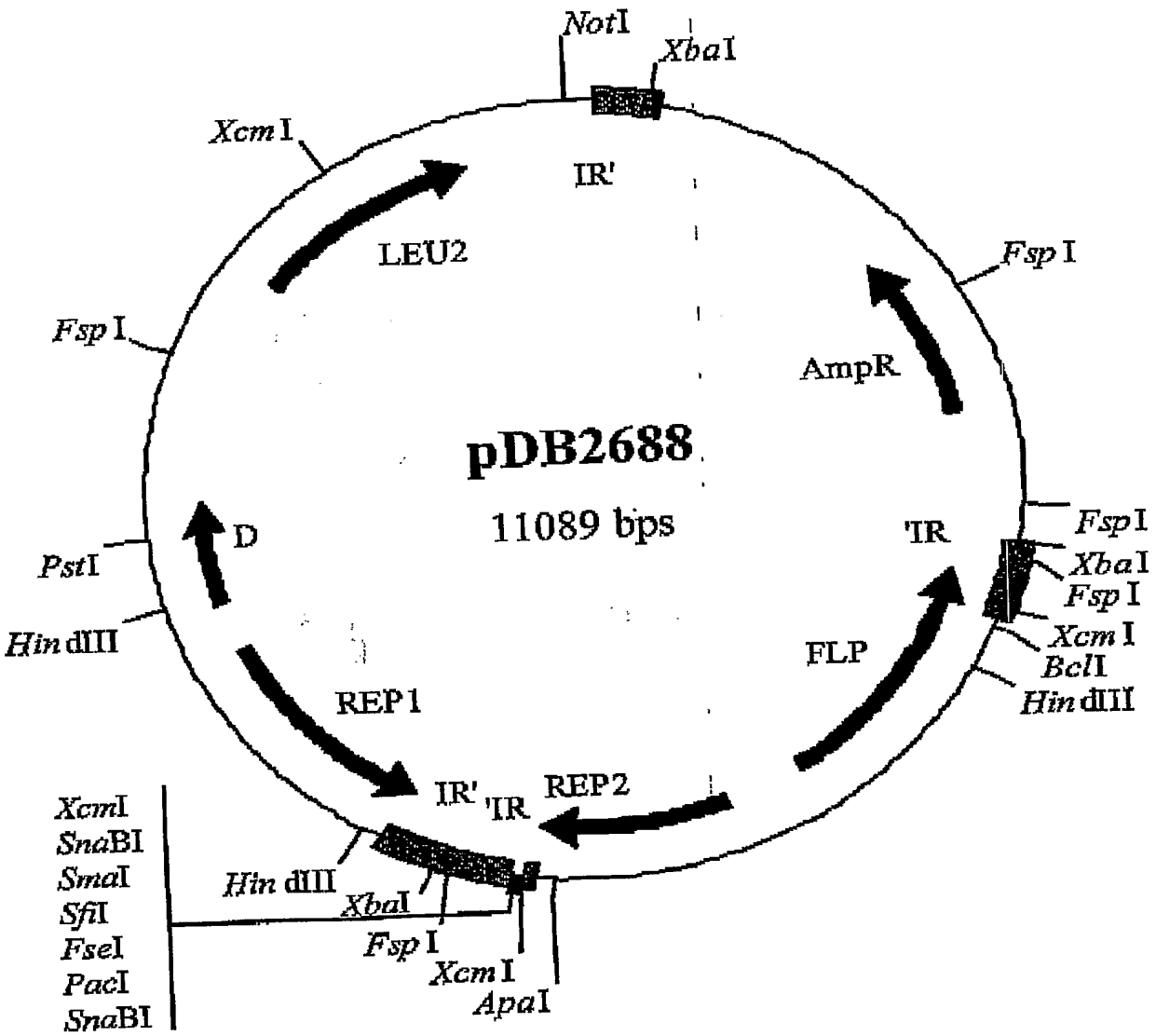
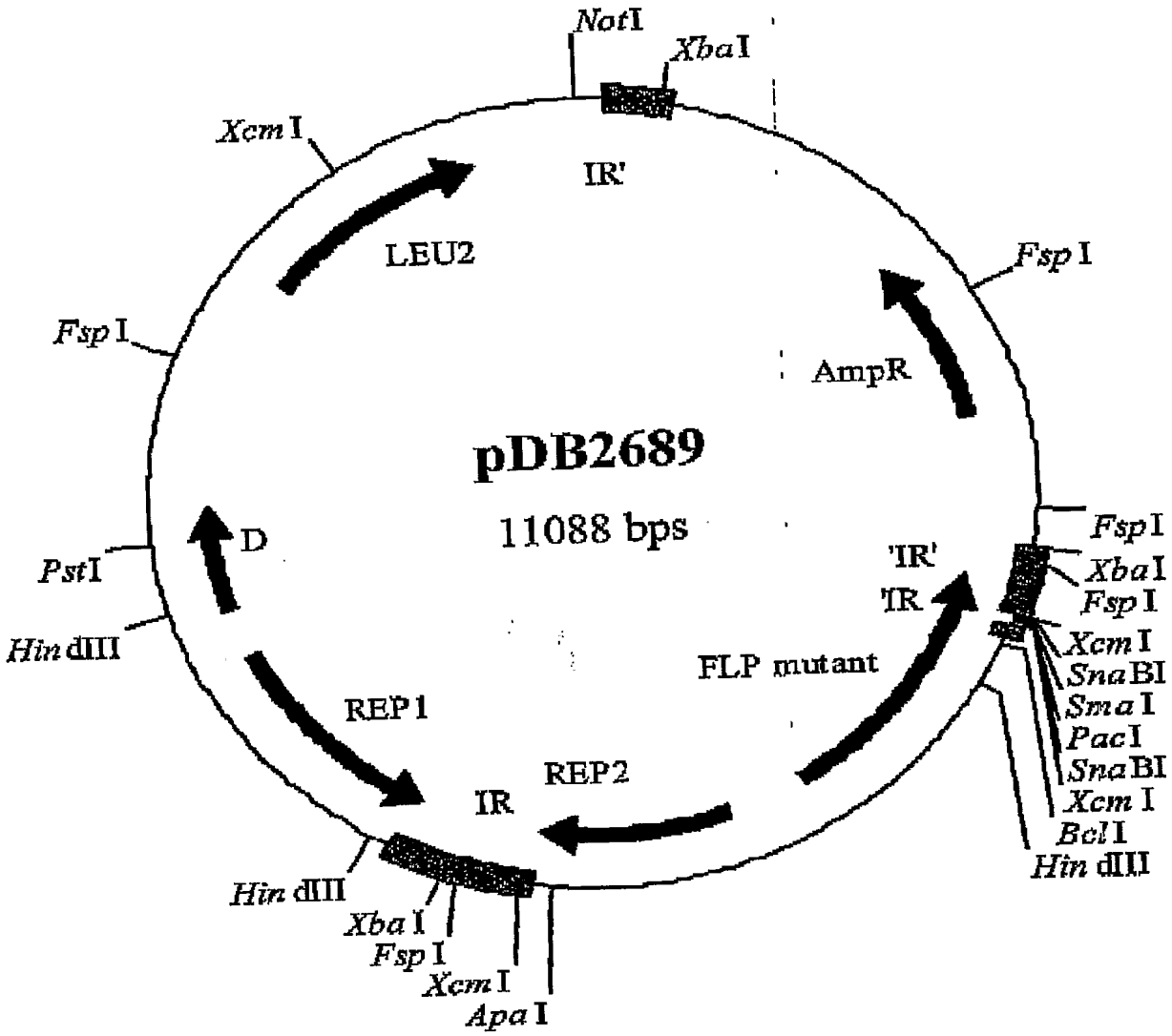


Figure 5



5

Figure 6

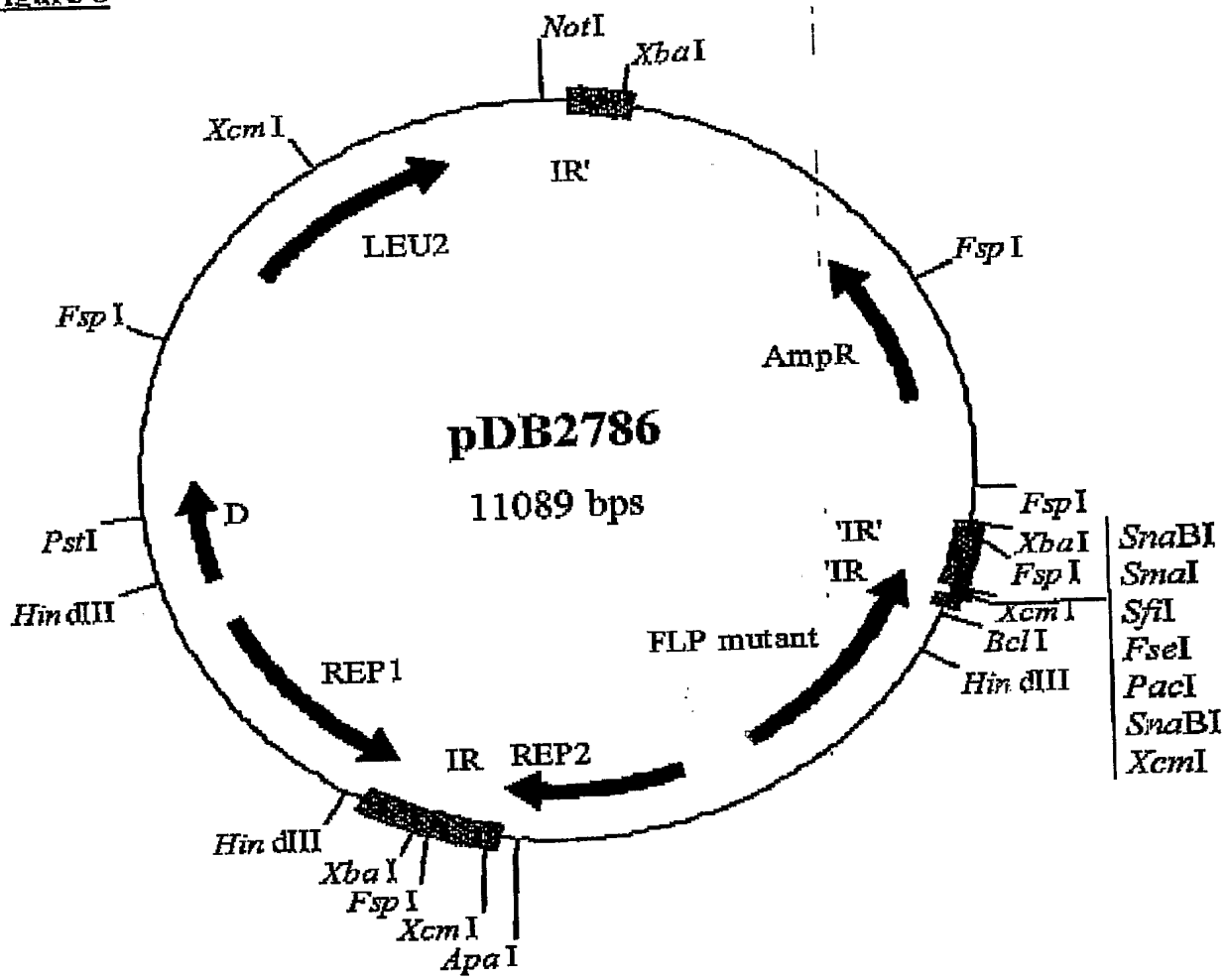


Figure 7

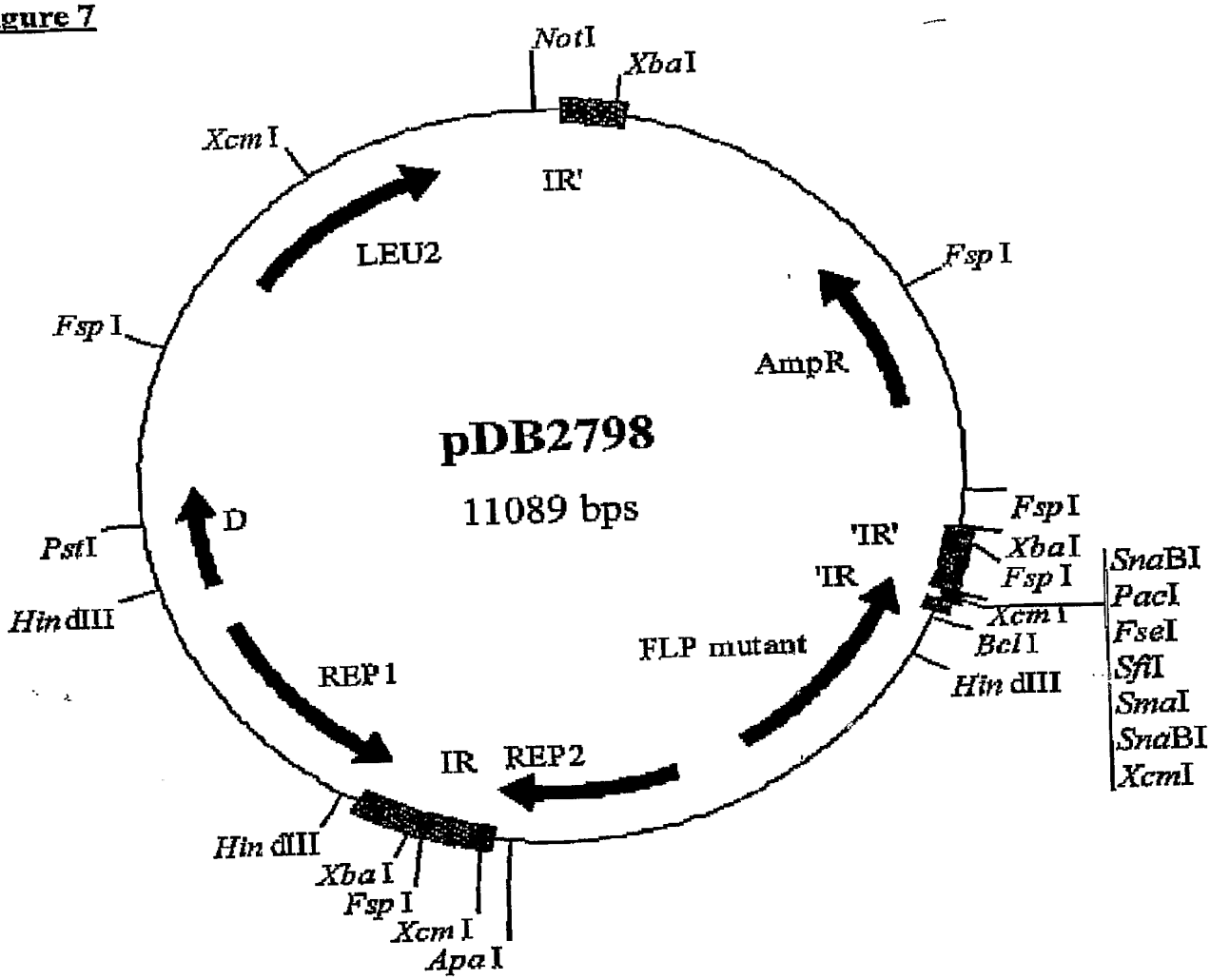
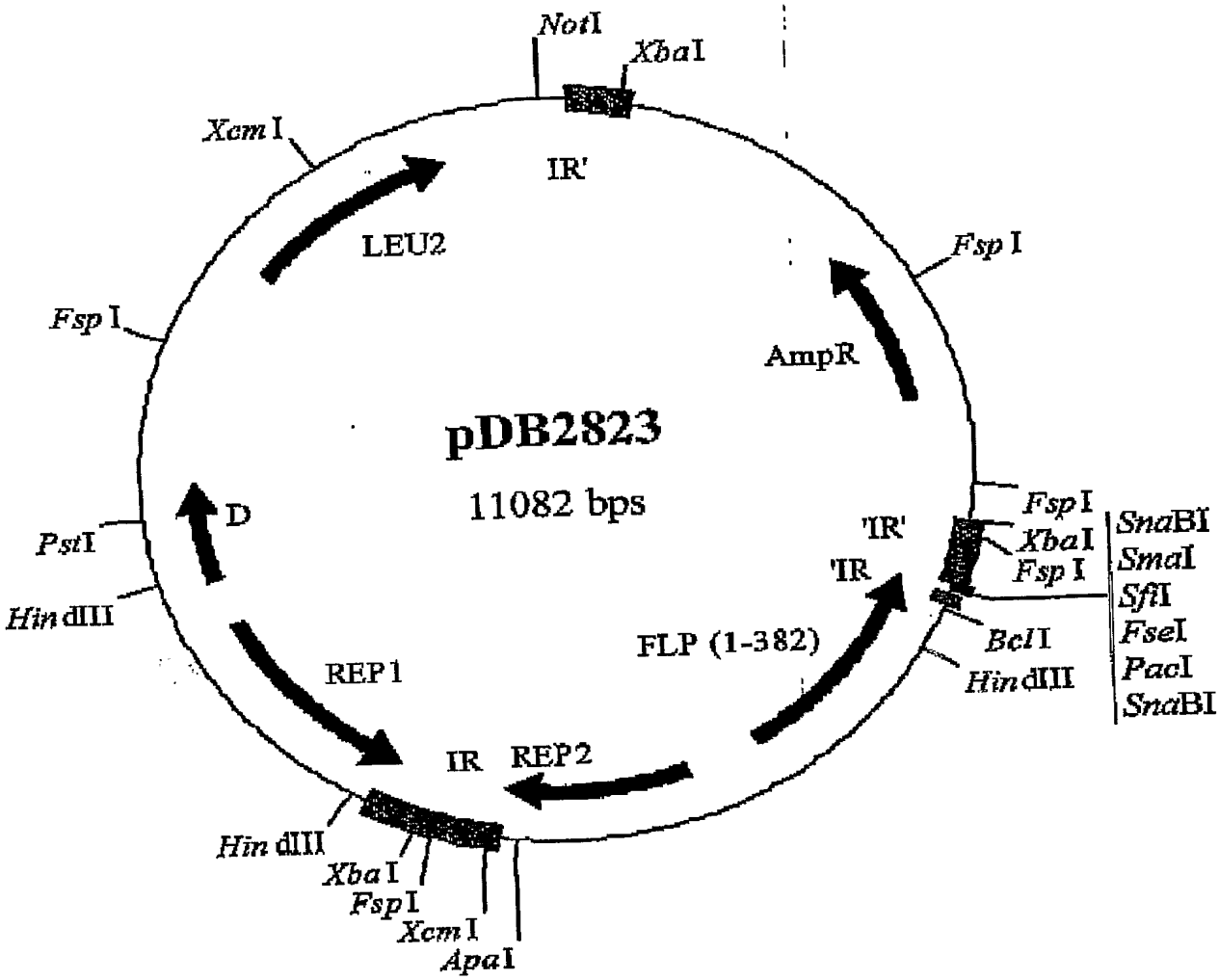


Figure 8



5

Figure 9

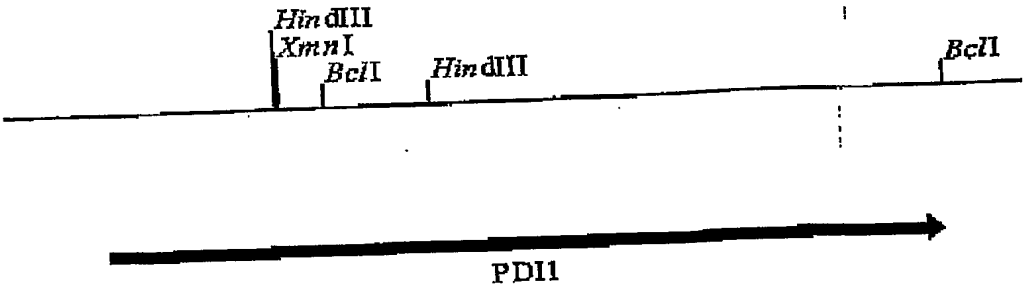


Figure 10

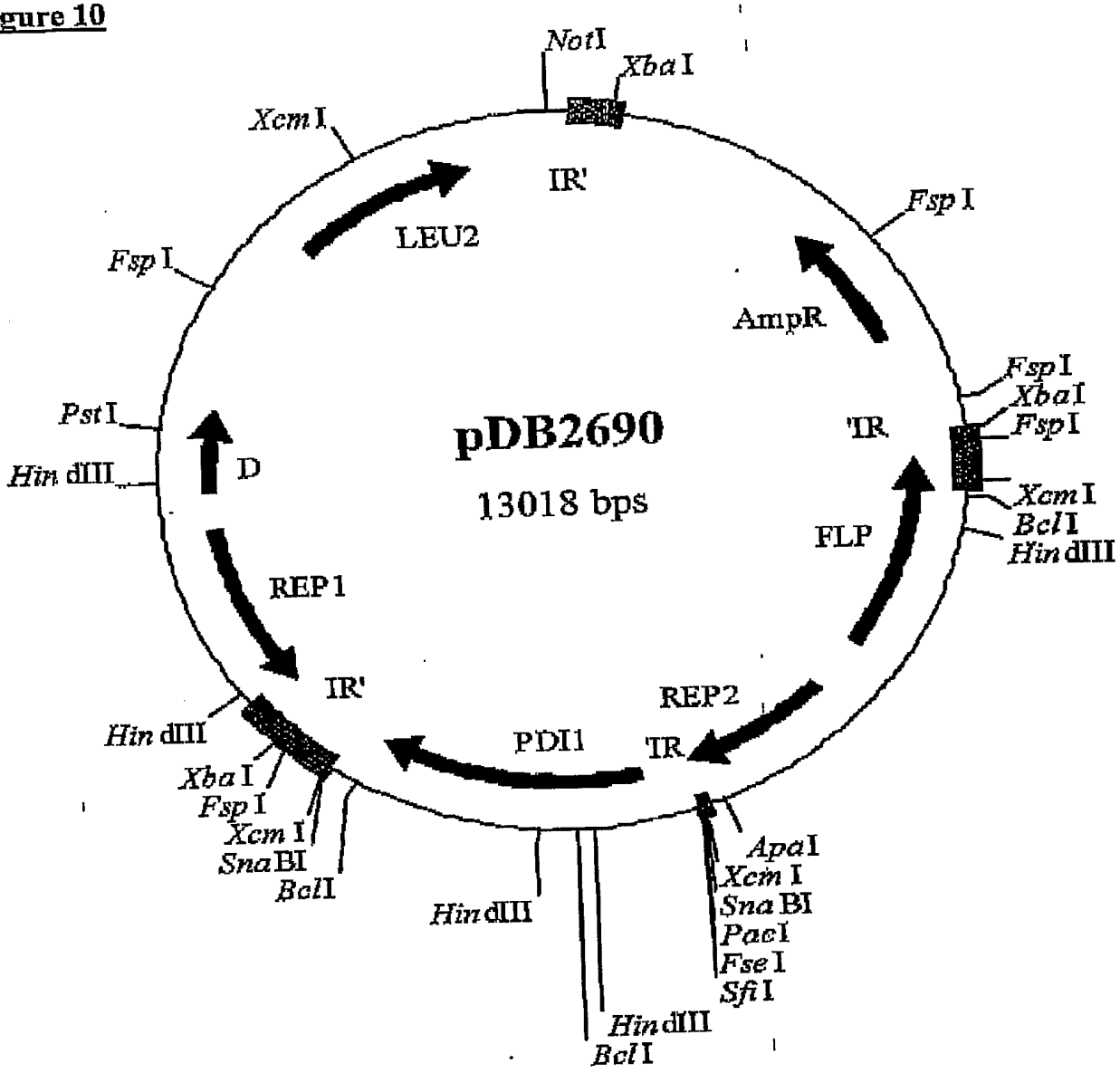
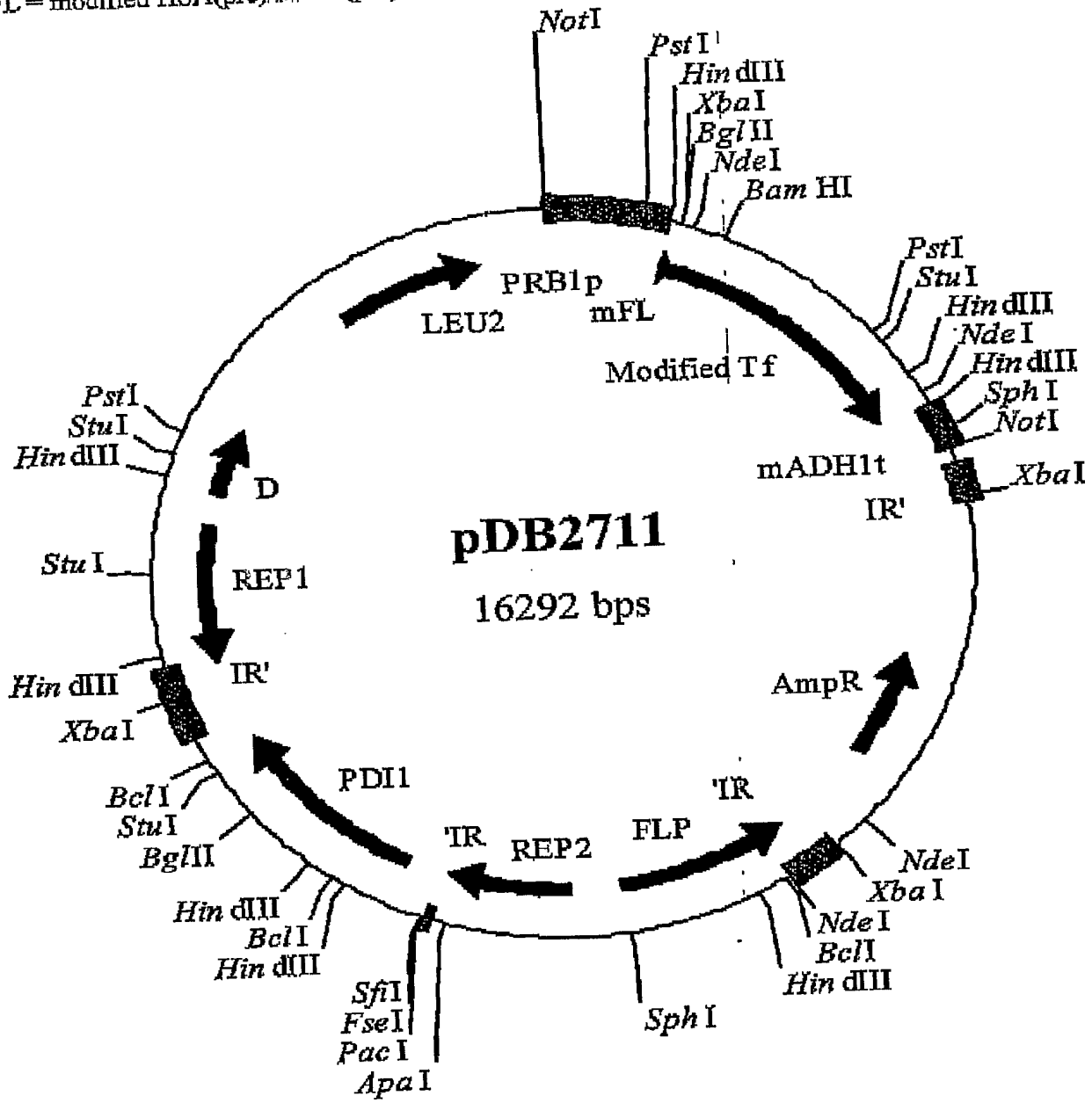


Figure 11

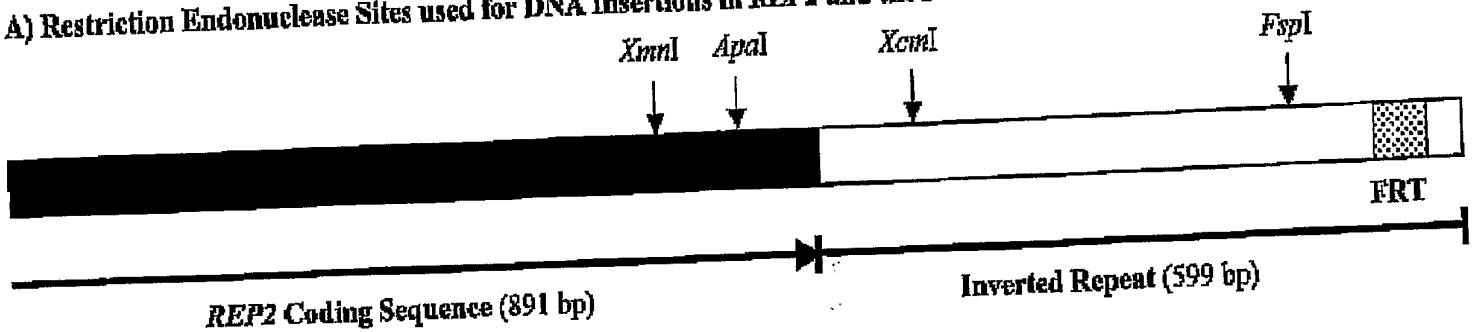
mFL = modified HSA(pre)/MFα1(pro) fusion leader sequence



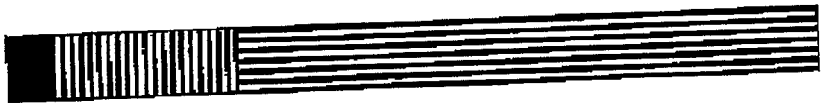
5

Figure 12

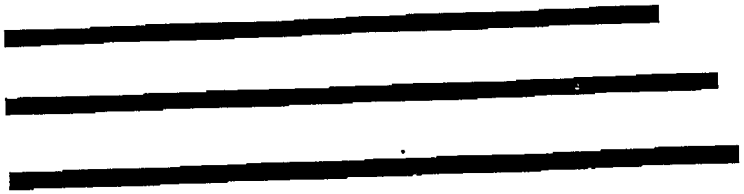
A) Restriction Endonuclease Sites used for DNA Insertions in *REP2* and the *REP2* Inverted Repeat



B) Rep2 Protein (296 amino acid residues)



C) Truncated Rep2 Protein Products



Rep2 (1-244) from termination at *Xmn*I-site

Rep2 (1-269) from termination at *Apa*I-site digested by T4 DNA polymerase

Rep2 (1-271) from termination at *Apa*I-site

Figure 13

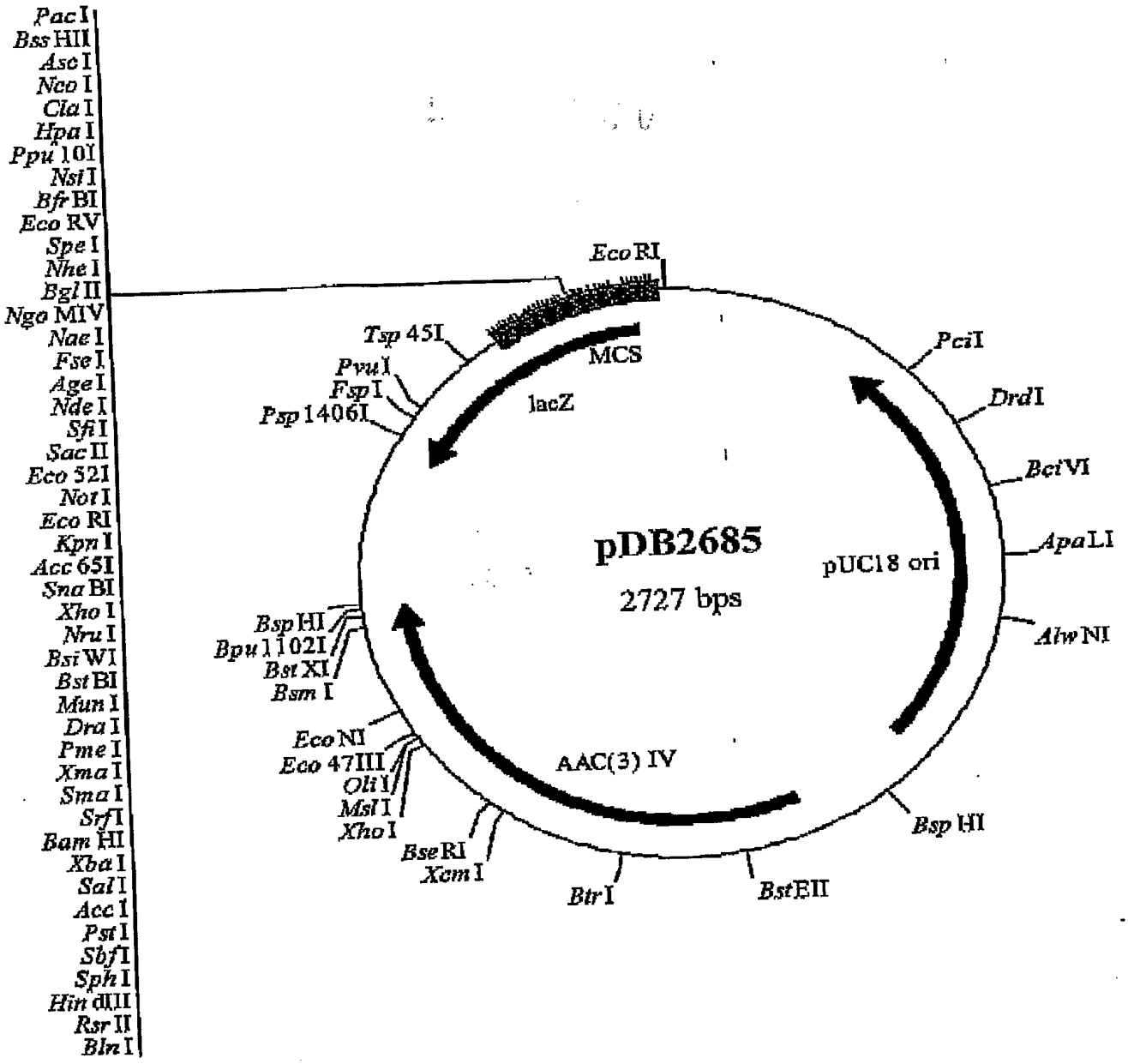


Figure 14

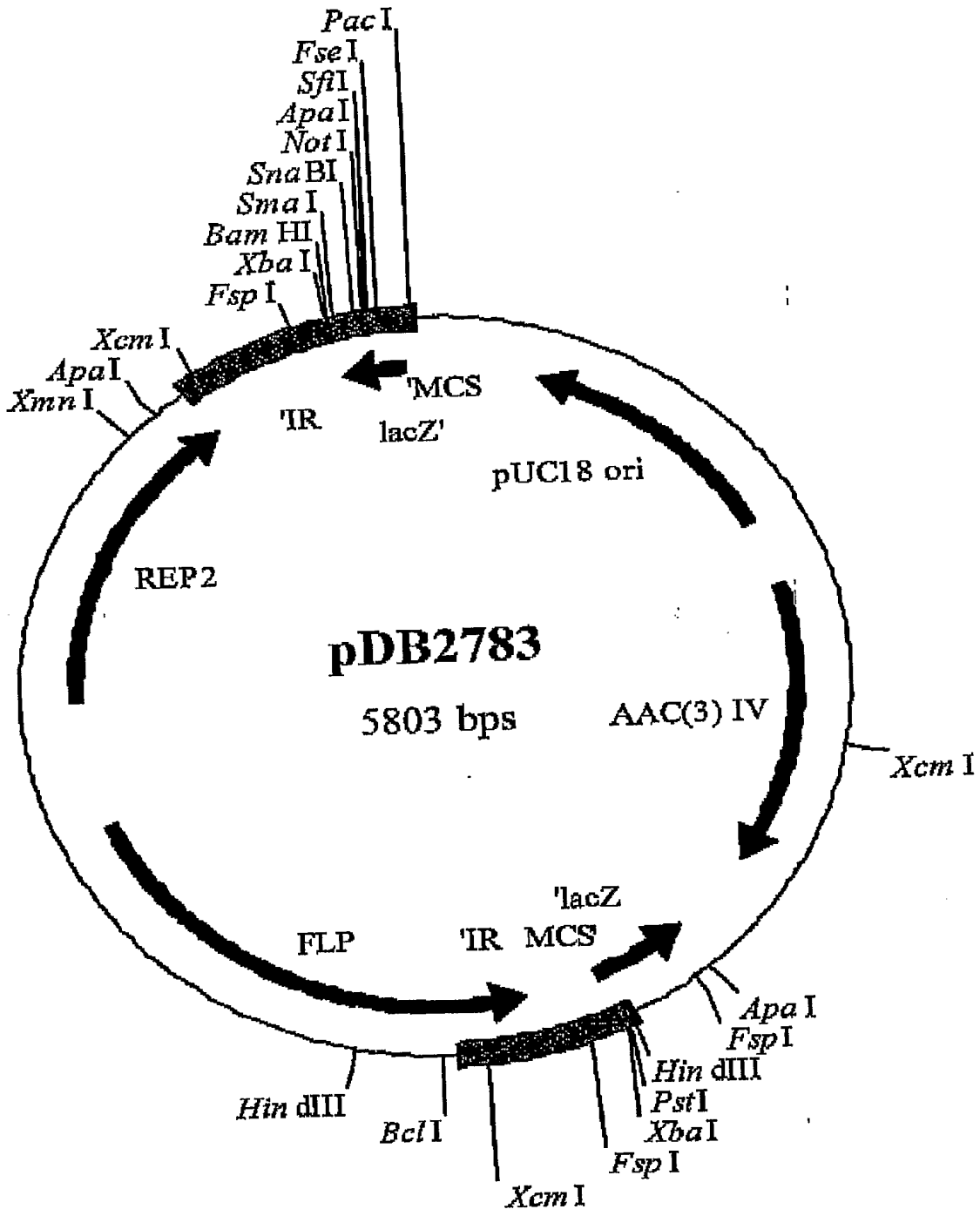


Figure 15

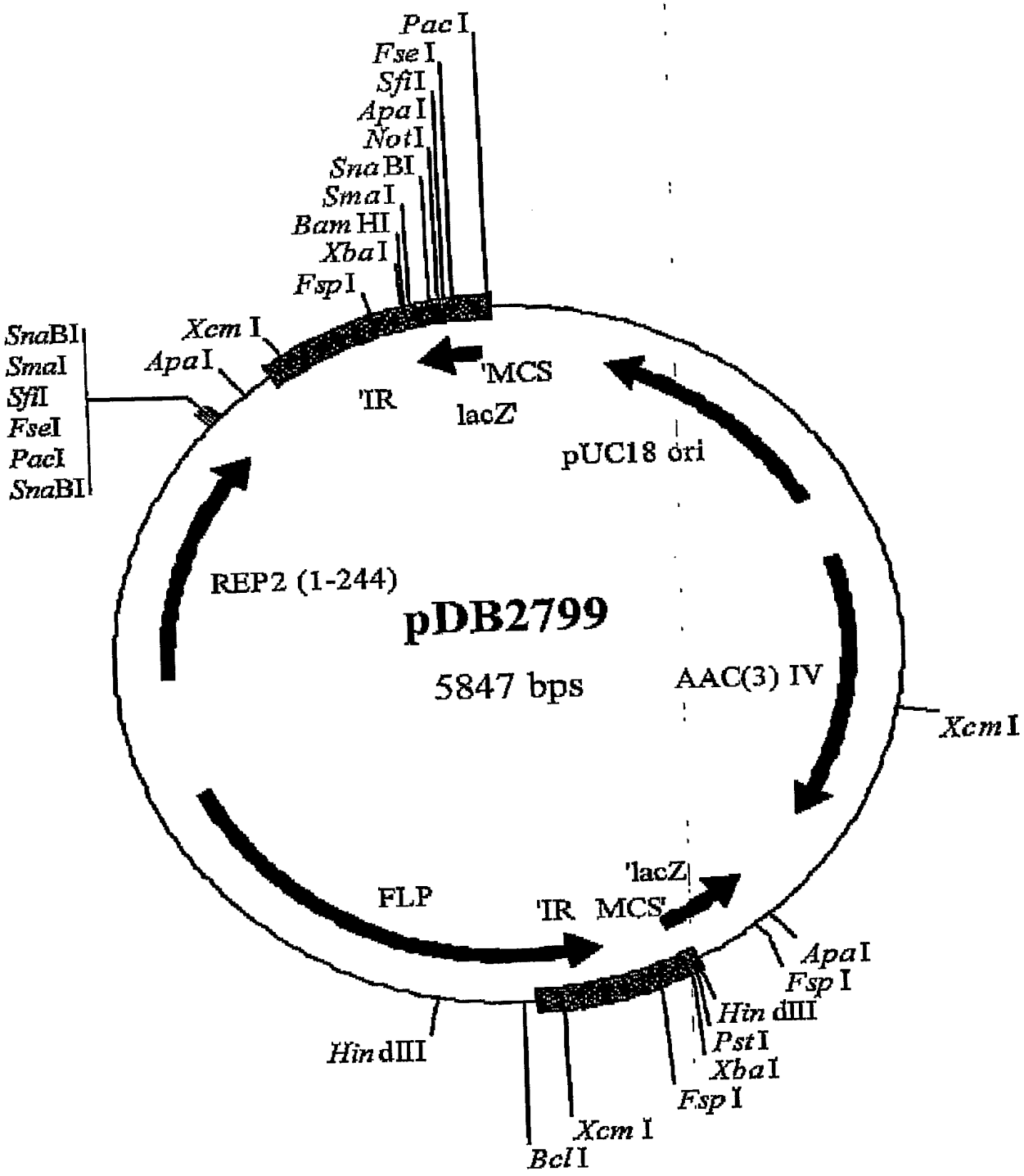


Figure 16

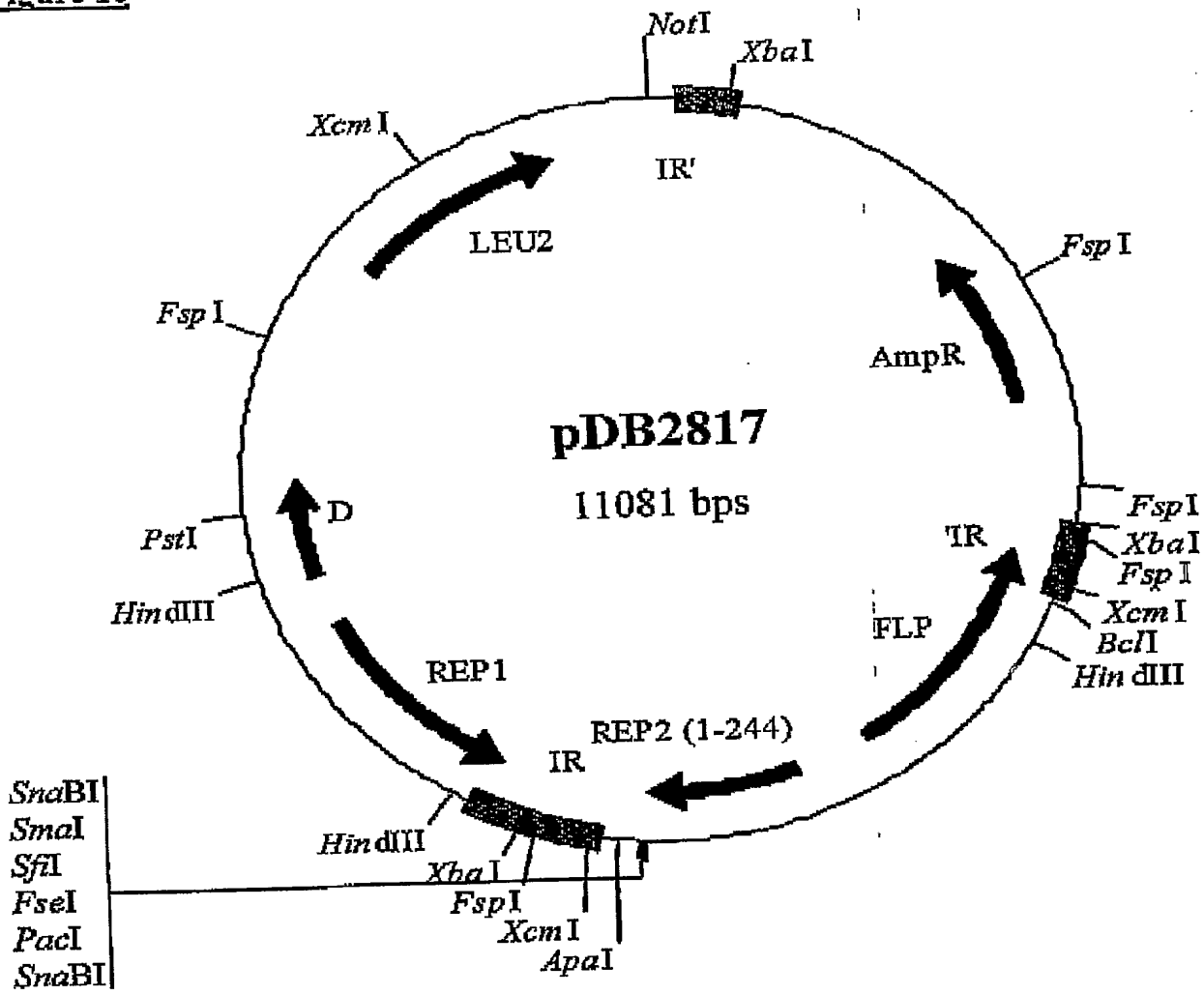


Figure 17

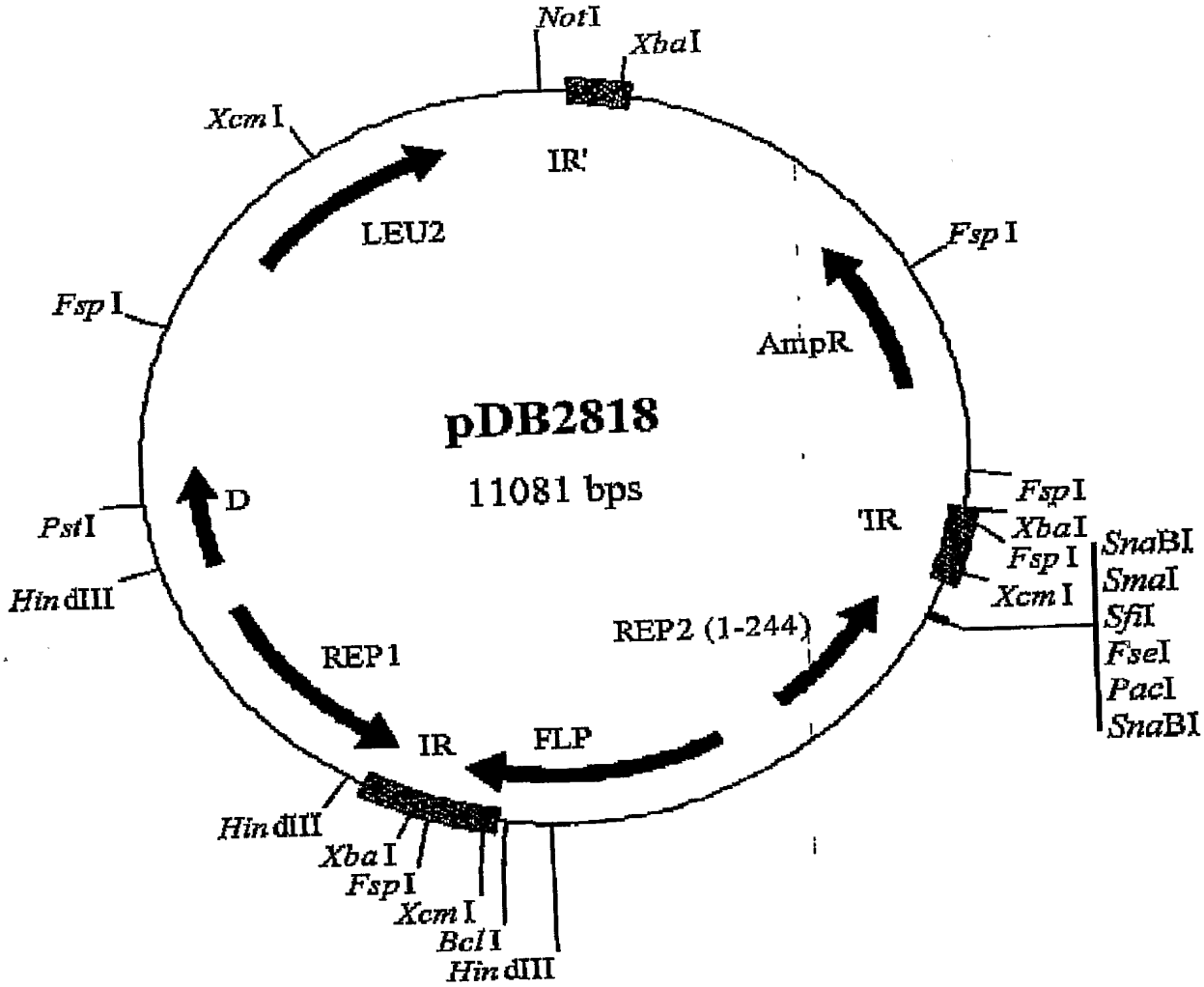


Figure 18

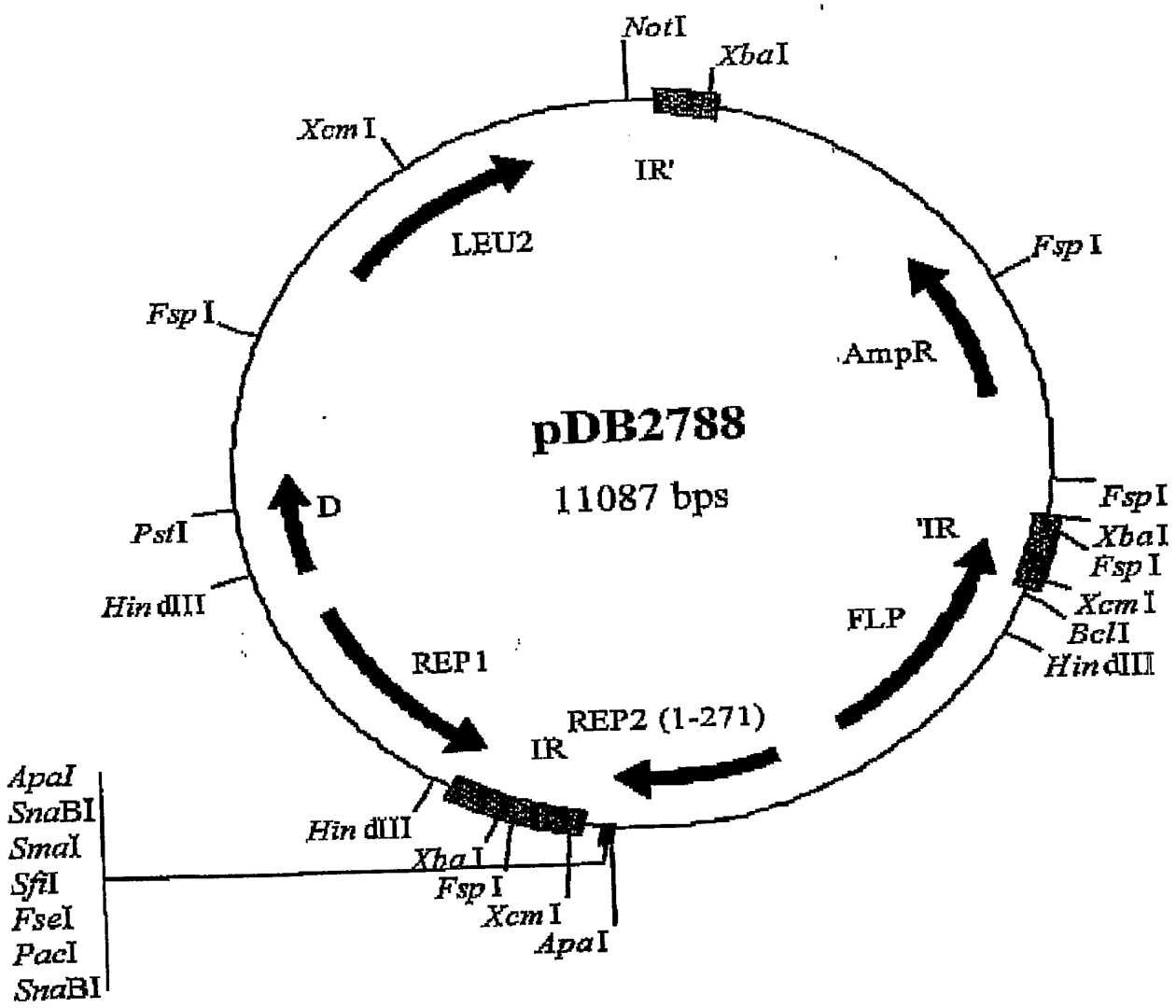


Figure 19

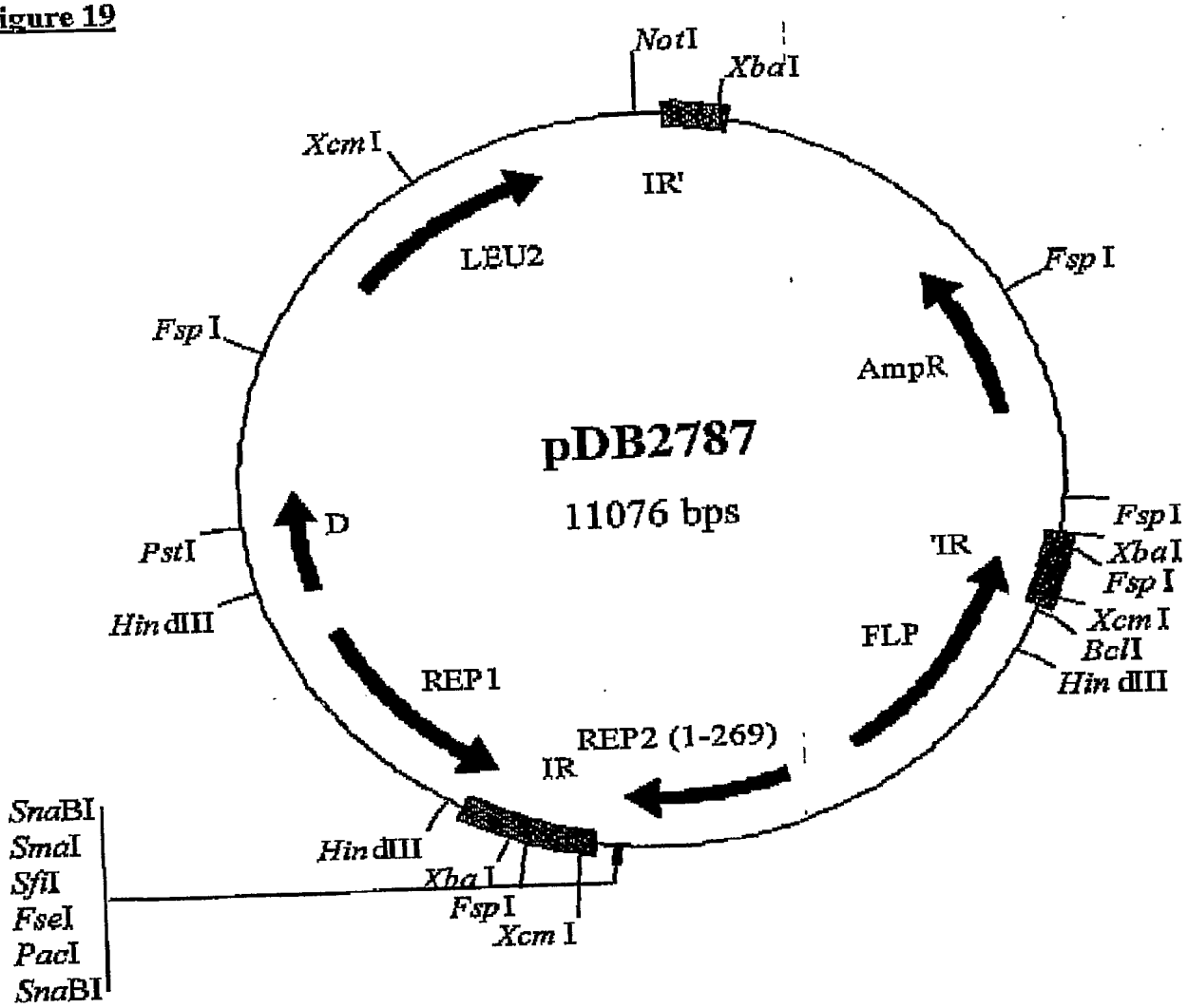


Figure 20

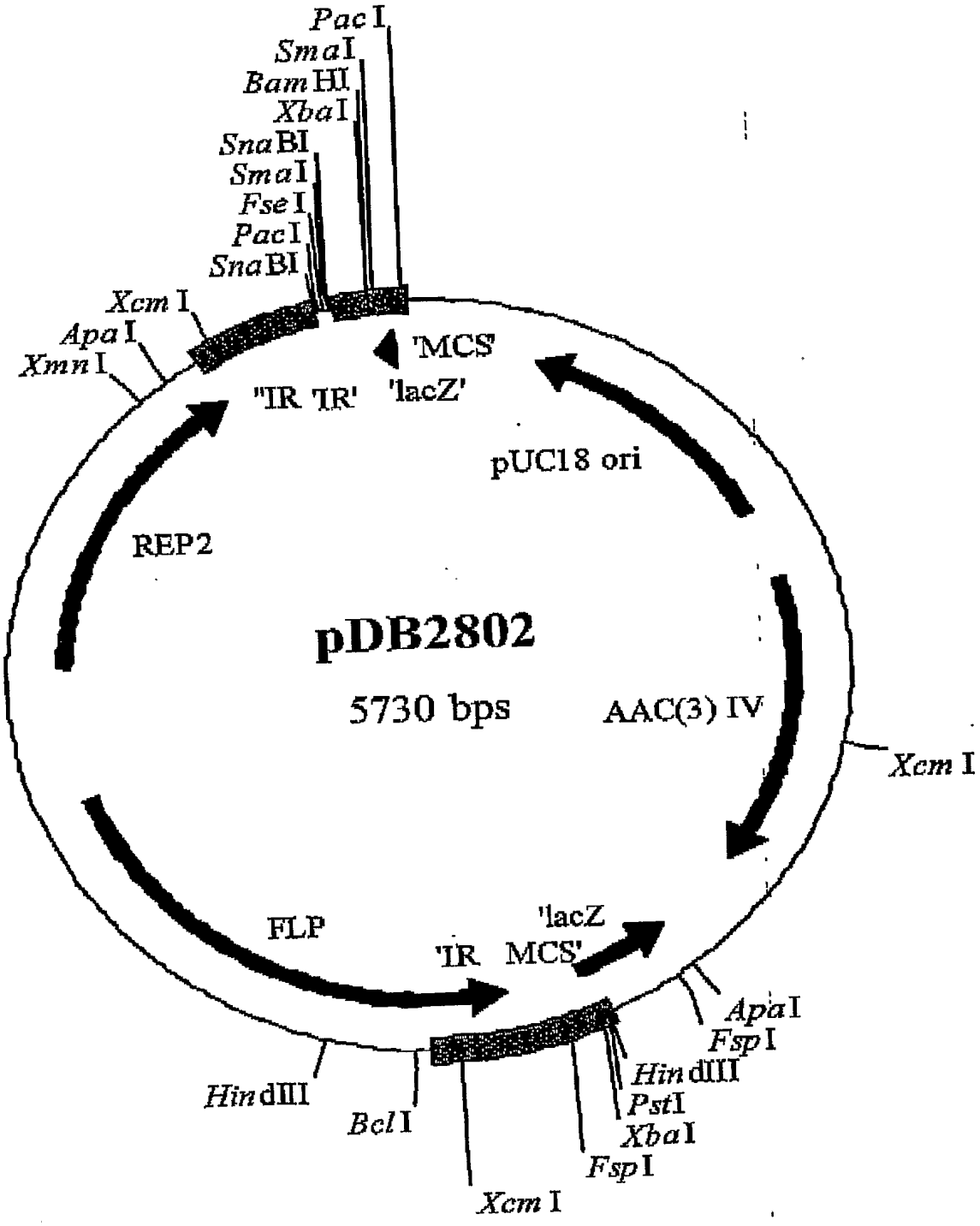


Figure 21

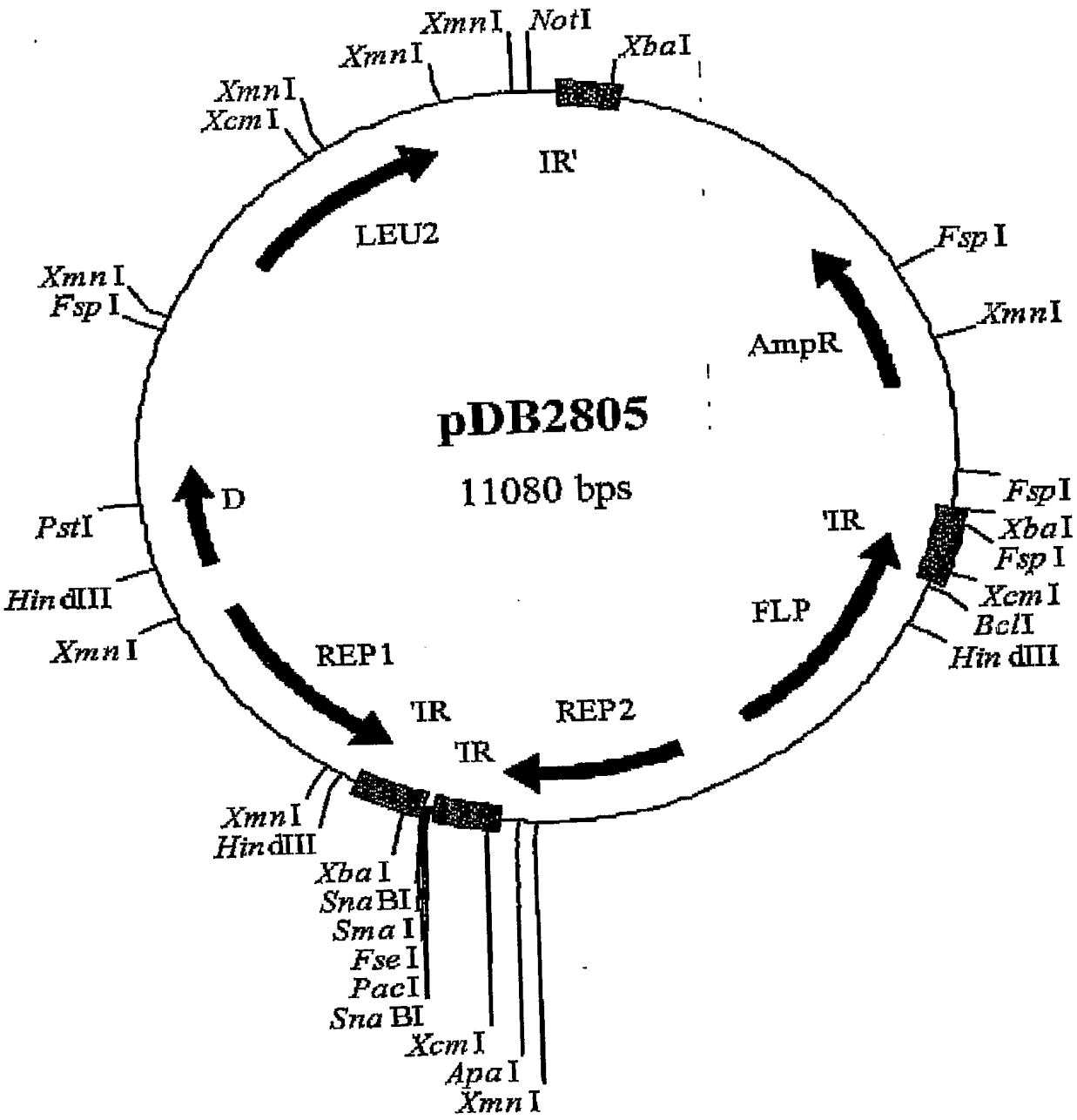


Figure 22

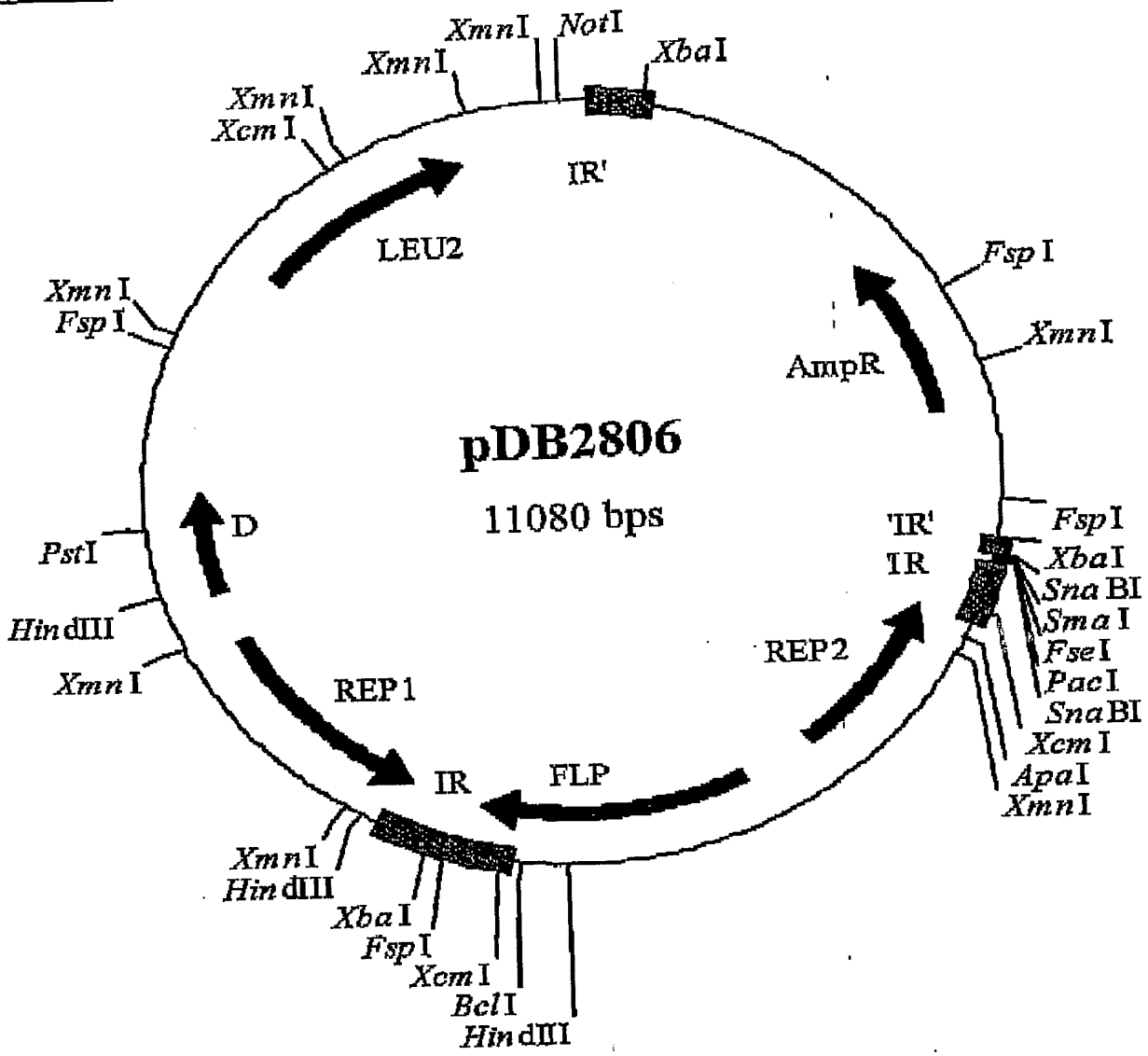


Figure 23

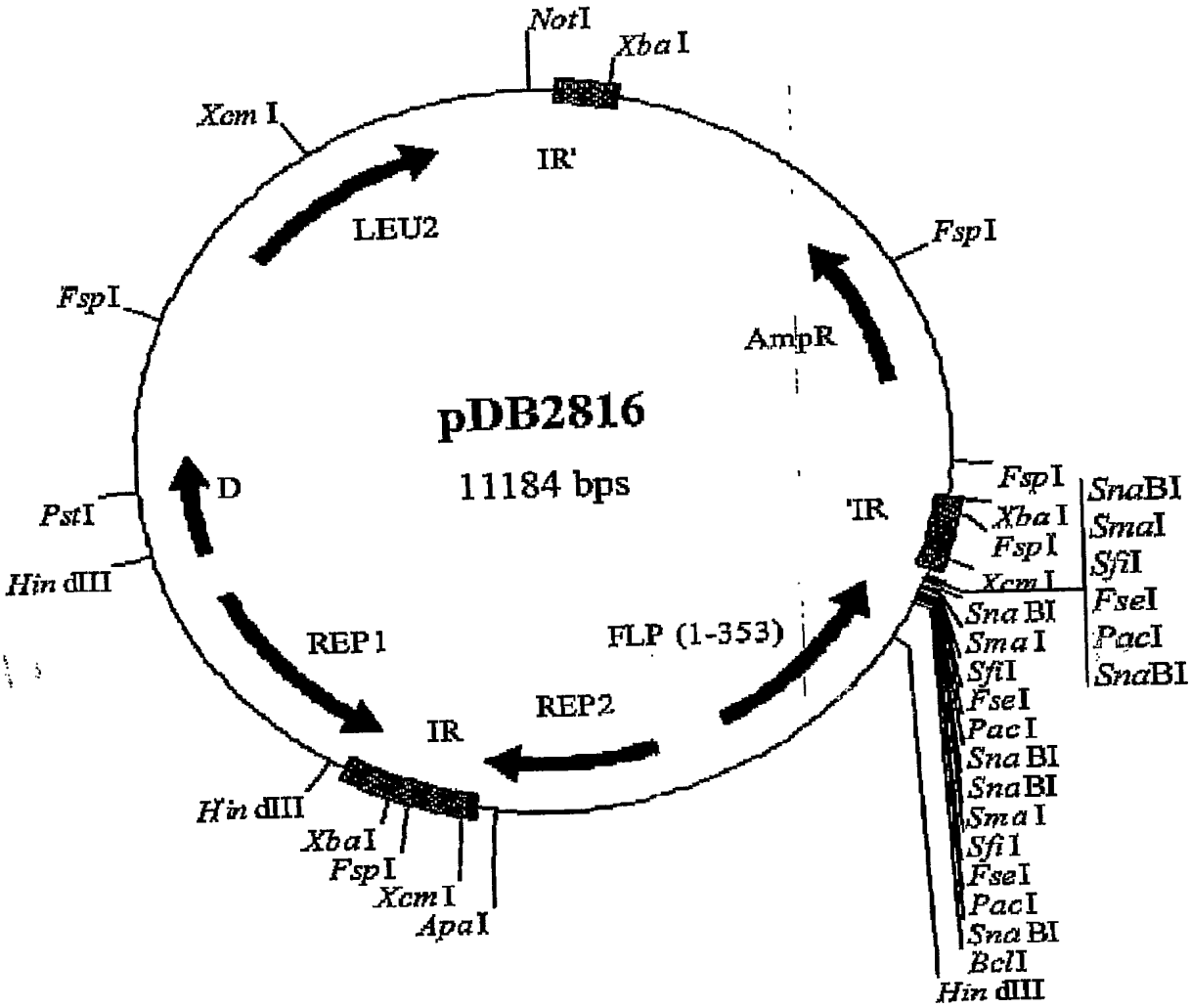


Figure 24

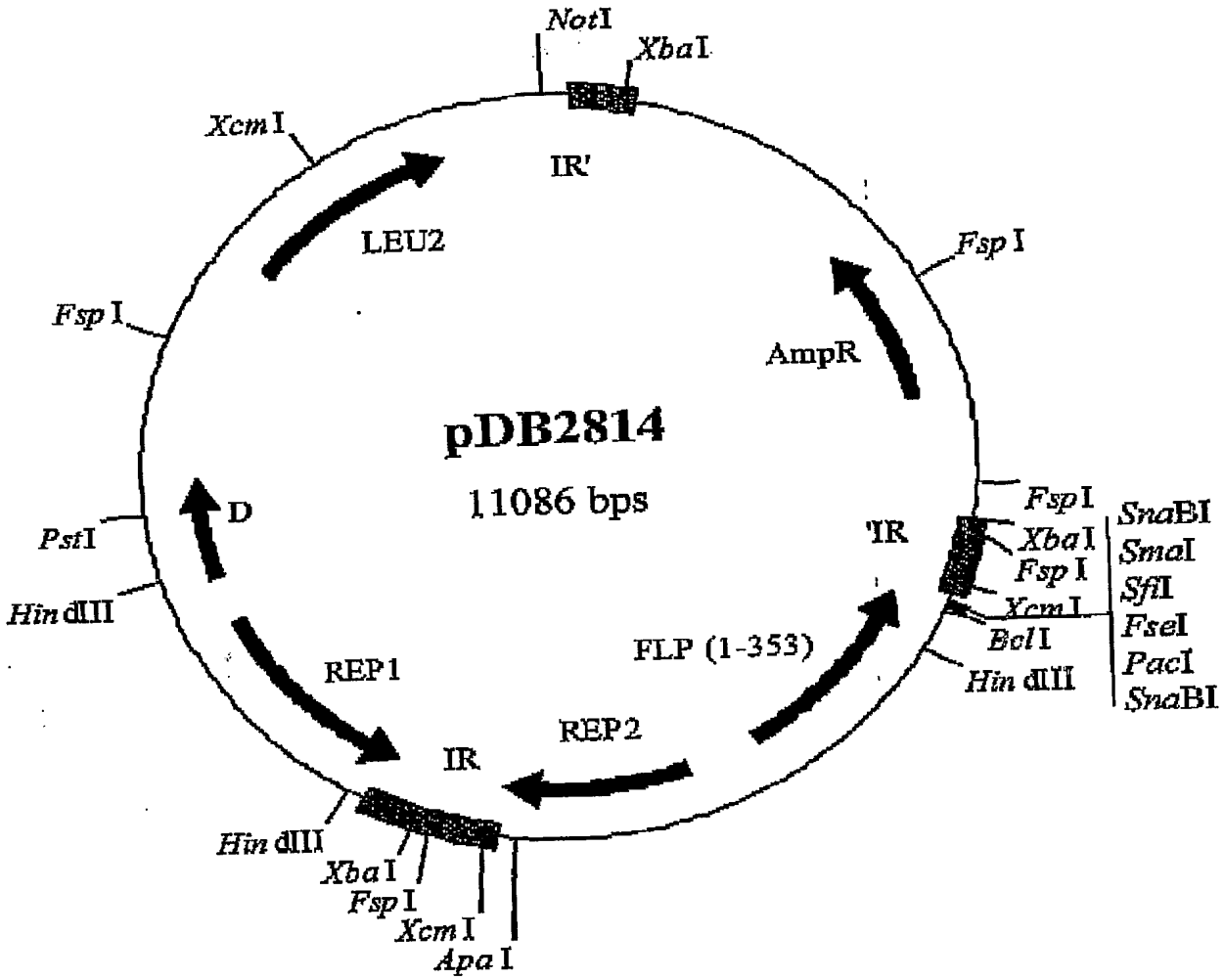


Figure 25

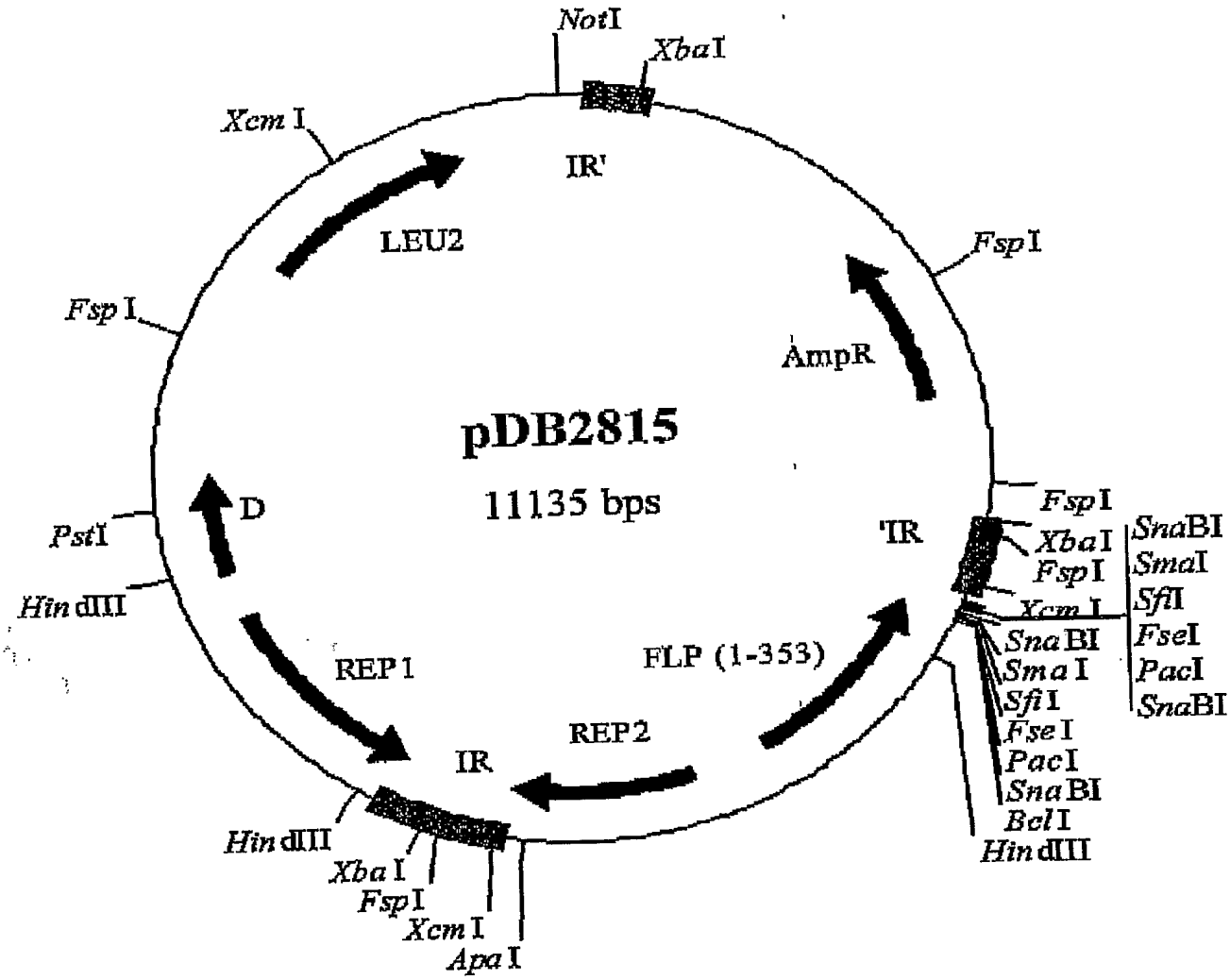


Figure 26

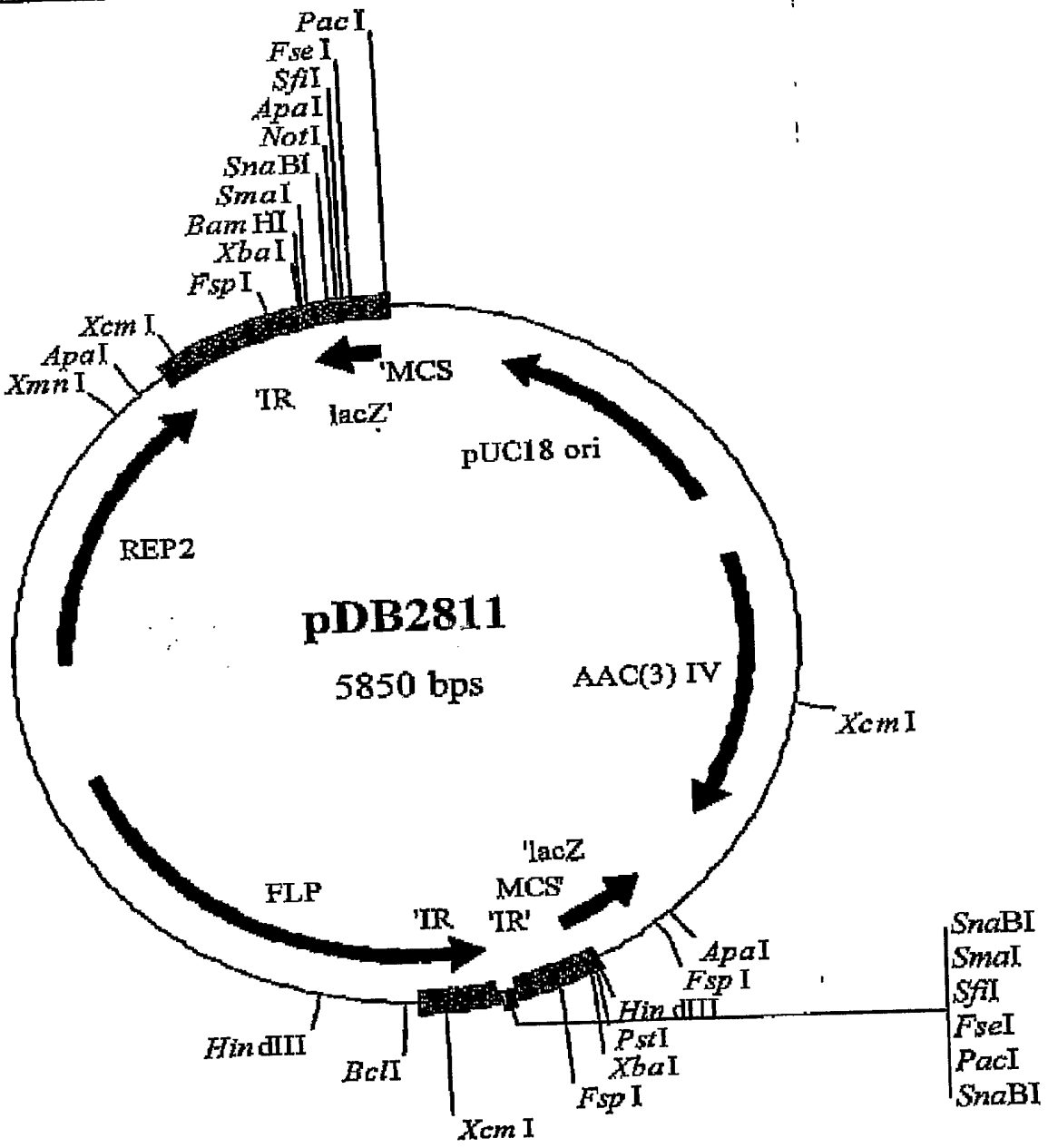


Figure 27

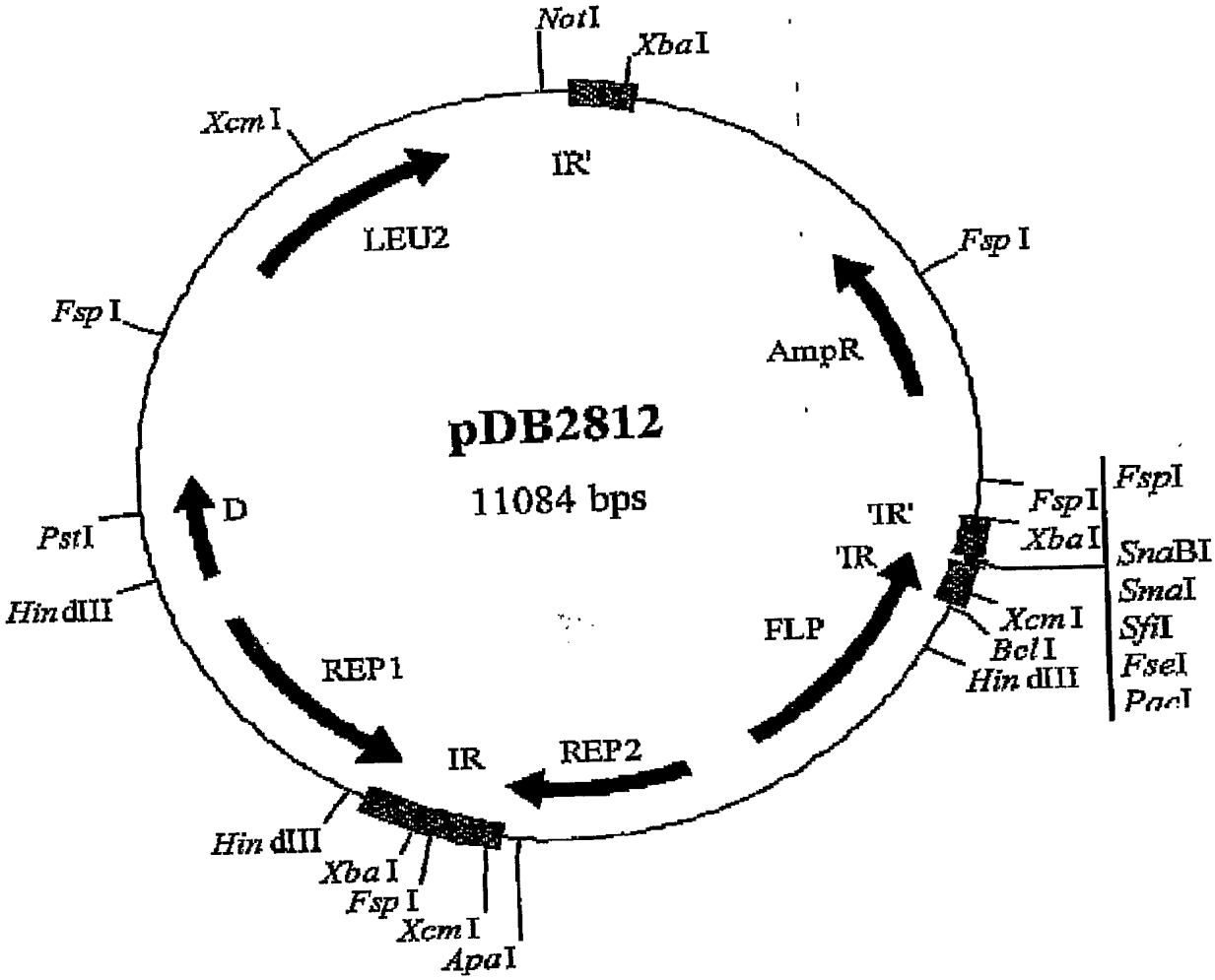

NotI
XbaI
XcmI
IR'
LEU2
FspI
FspI
AmpR
pDB2812
11084 bps
FspI
XbaI
SnaBI
SmaI
SfiI
FseI
PacI
IR'
IR
D
PstI
HindIII
FLP
XcmI
BclI
HindIII
REP1
IR
REP2
HindIII
XbaI
FspI
XcmI
ApaI

Figure 28

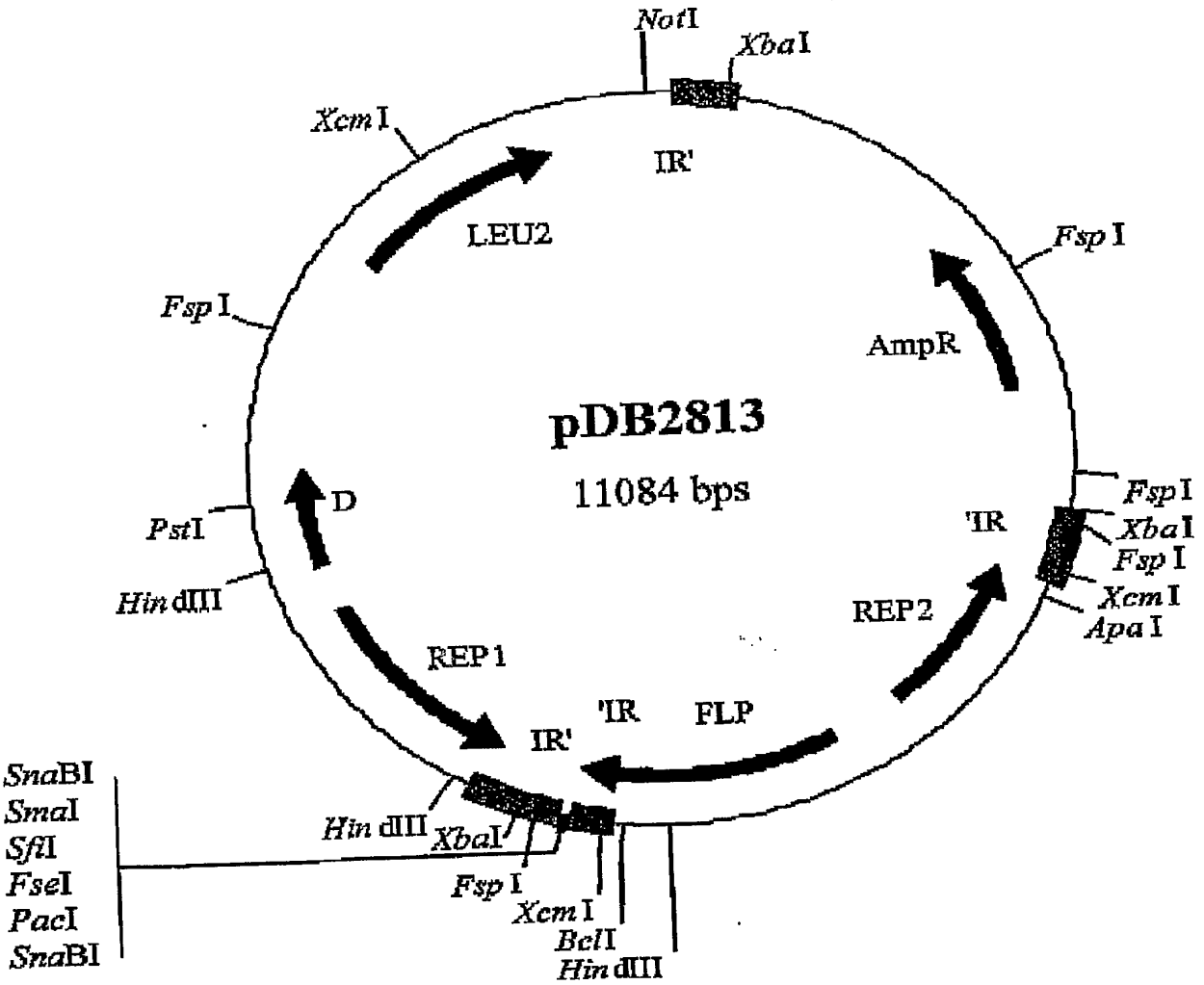


Figure 29

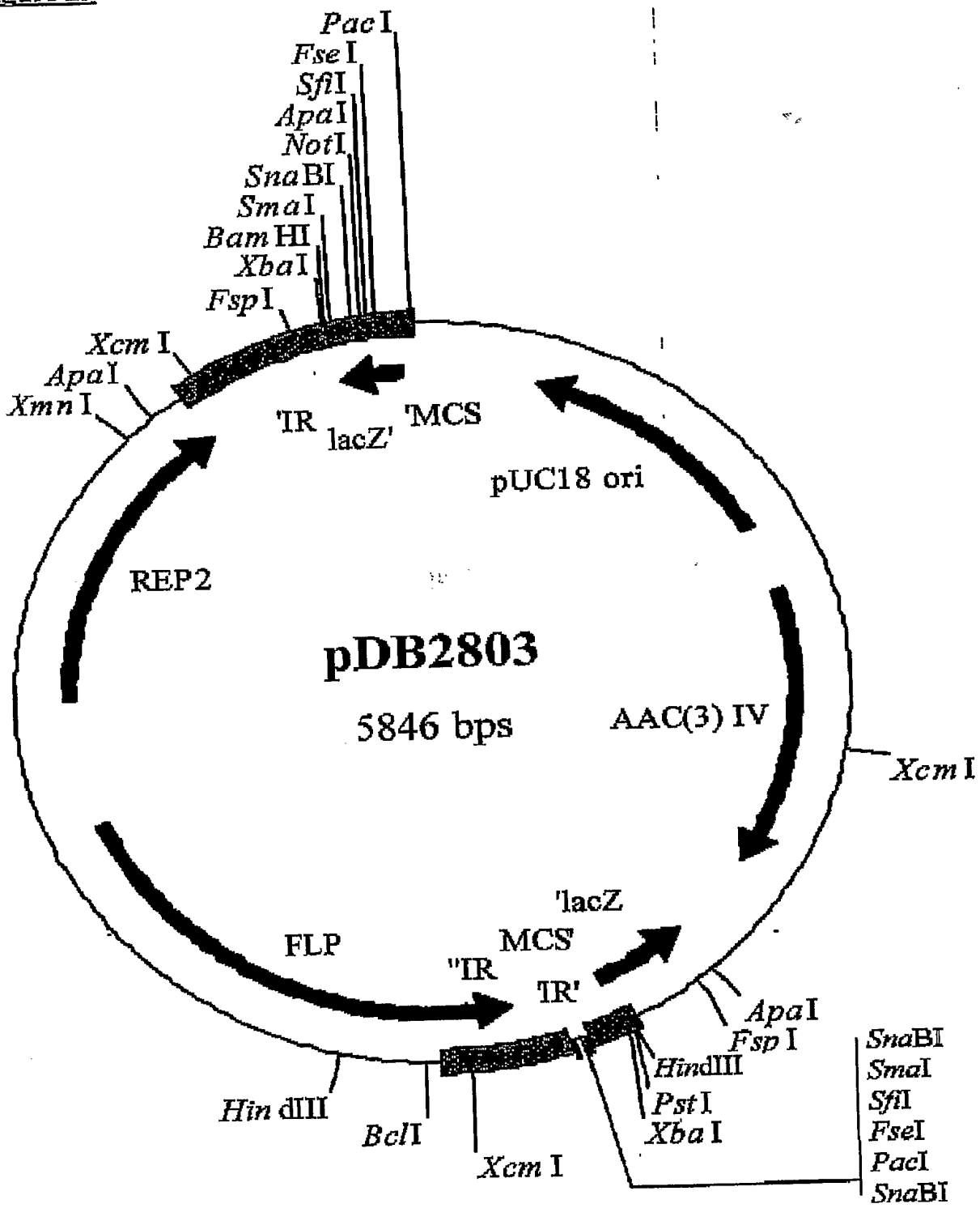


Figure 30

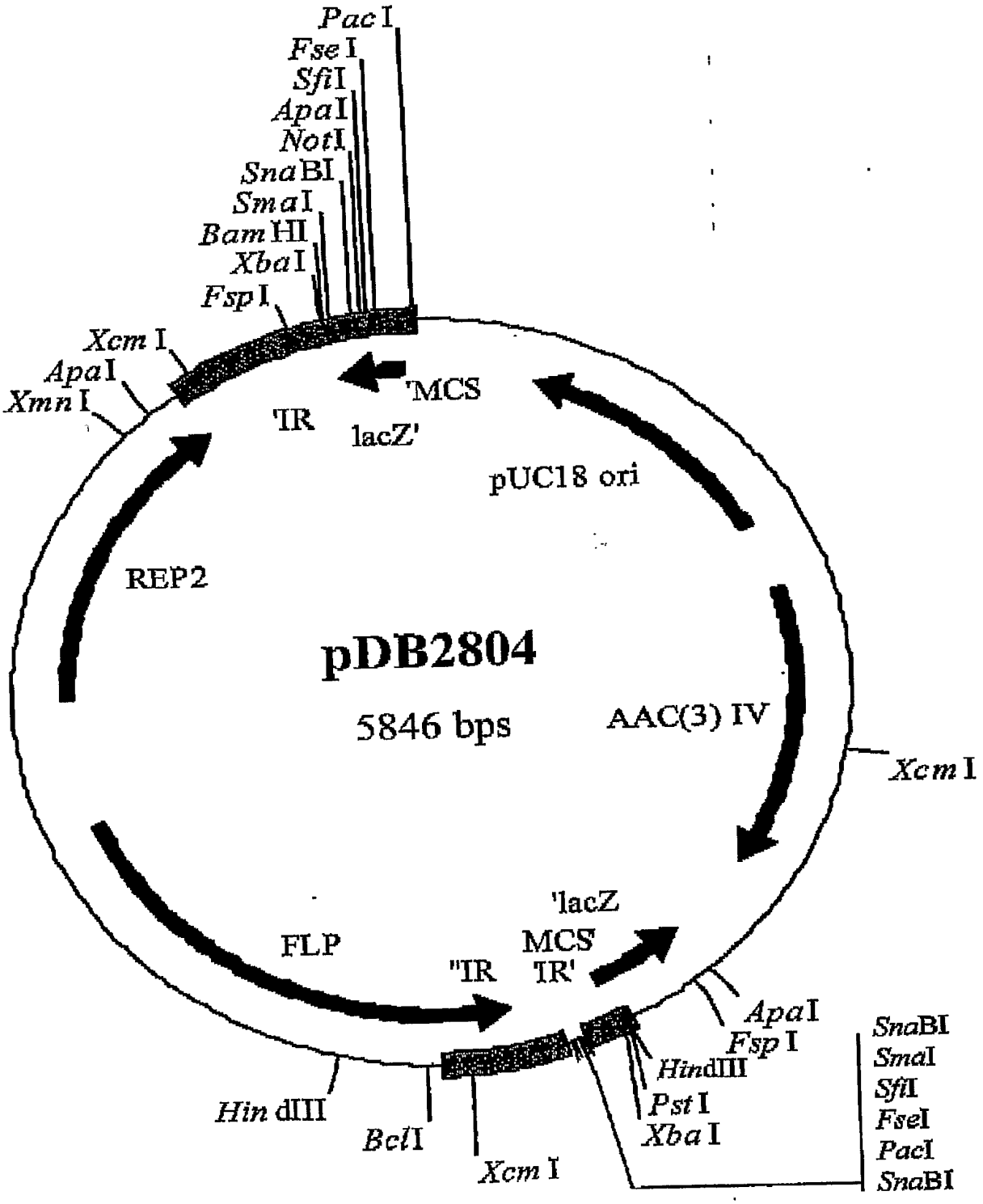


Figure 31

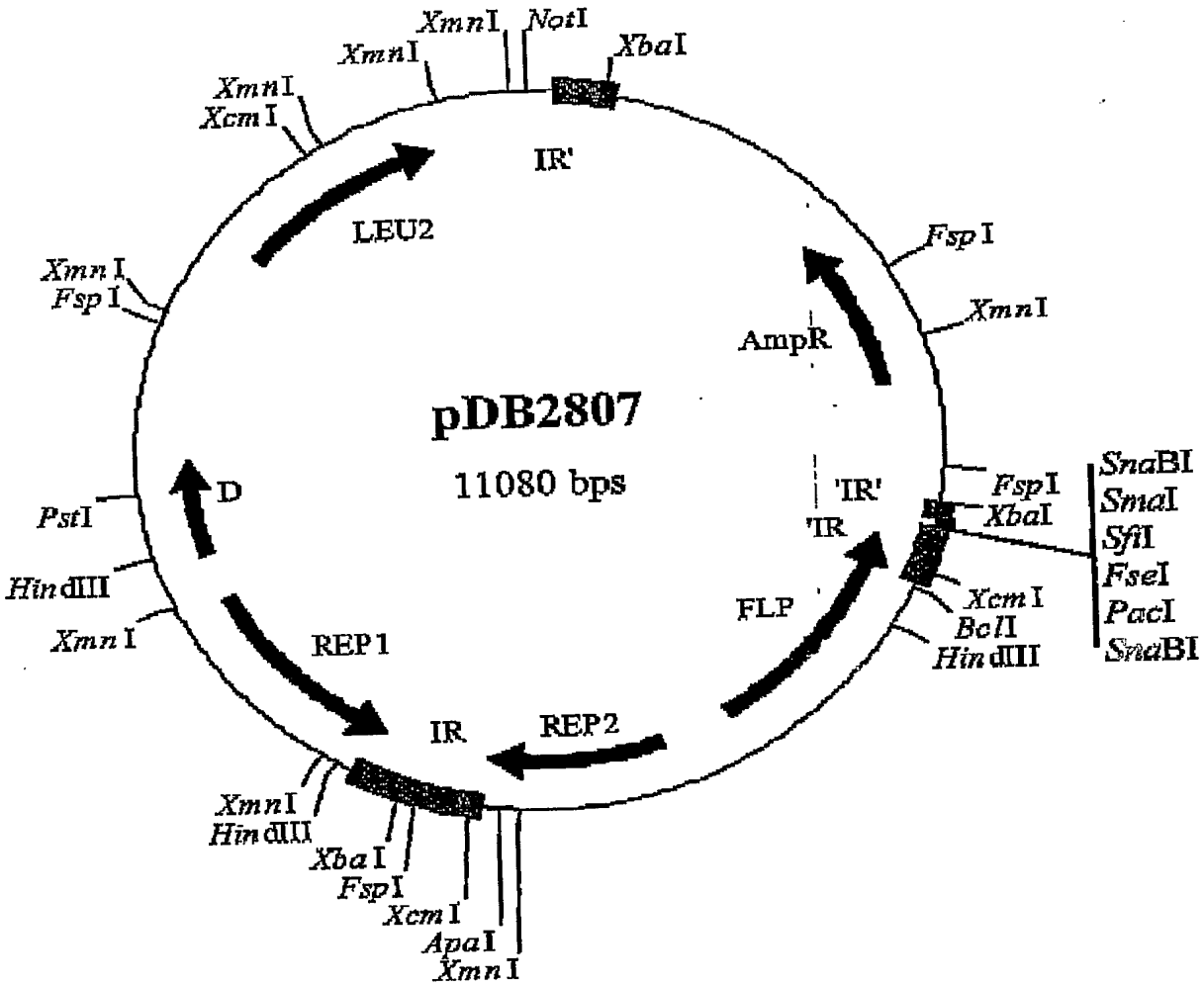


Figure 32

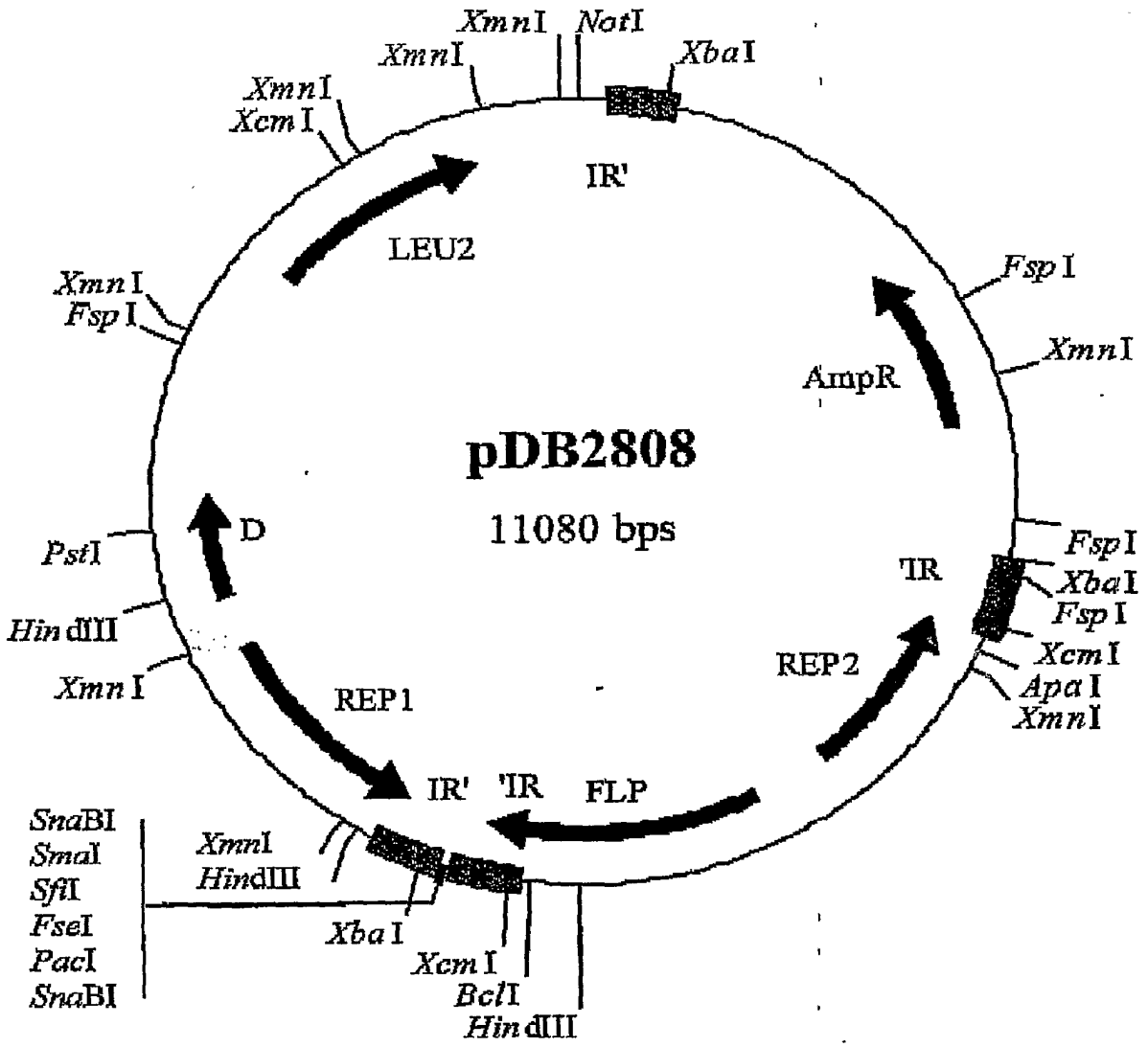


Figure 33A

Table 3 (part 1)

Plasmid	Insertion Site	Insertion Site Details	2μm Form	Initial OD_{600}	Leucine Prototrophs		
					Number	%	Average
pSAC35	-	-	B	0.074	100/100	100%	99%
pSAC35	-	-	B	0.068	95/98	97%	
pDB2818	*Xmn*I	REP2 (1-244)	A	0.062	32/100	32%	42%
pDB2818	*Xmn*I	REP2 (1-244)	A	0.059	52/100	52%	
pDB2787	*Apa*I/T4 pol.	REP2 (1-269)	B	0.054	34/100	34%	45%
pDB2787	*Apa*I/T4 pol.	REP2 (1-269)	B	0.084	45/100	45%	
pDB2788	*Apa*I	REP2 (1-271)	B	0.066	23/100	23%	33%
pDB2788	*Apa*I	REP2 (1-271)	B	0.051	43/100	43%	

Figure 33B

Table 3 (part 2)

Plasmid	Insertion Site	Insertion Site Details	2μm Form	Initial OD_{600}	Leucine Prototrophs		
					Number	%	Average
pDB2688	*Xcm*I	Inverted Repeat	B	0.055	100/100	100%	100%
PDB2688	*Xcm*I	Inverted Repeat	B	0.066	100/100	100%	
pDB2806	*Fsp*I	Inverted Repeat	A	0.073	100/100	100%	100%
pDB2806	*Fsp*I	Inverted Repeat	A	0.070	100/100	100%	
pDB2817	*Xmn*I	REP2 (1-244)	B	0.063	36/100	36%	35%
pDB2817	*Xmn*I	REP2 (1-244)	B	0.082	34/100	34%	
pDB2805	*Fsp*I	Inverted Repeat	B	0.069	100/100	100%	100%
pDB2805	*Fsp*I	Inverted Repeat	B	0.078	100/100	100%	
pDB2814	*Bcl*I	FLP (1-353), 1× Insert	B	0.080	69/100	69%	67%
pDB2814	*Bcl*I	FLP (1-353), 1× Insert	B	0.057	64/100	64%	
pDB2815	*Bcl*I	FLP (1-353), 2× Insert	B	0.067	70/100	70%	67%
pDB2815	*Bcl*I	FLP (1-353), 2× Insert	B	0.068	64/100	64%	

Figure 33C

Table 3 (part 3)

Plasmid	Insertion Site	Insertion Site Details	2µm Form	Initial OD_{600}	Leucine Prototrophs		
					Number	%	Average
pDB2816	*Bcl*I	FLP (1-353), 3× Insert	B	0.069	67/100	67%	74%
pDB2816	*Bcl*I	FLP (1-353), 3× Insert	B	0.056	81/100	81%	
pDB2689	*Xcm*I	C-terminal FLP Mutant (FLP 1-384, plus 56 other residues)	B	0.054	73/100	73%	75%
pDB2689	*Xcm*I	C-terminal FLP Mutant (FLP 1-384, plus 56 other residues)	B	0.056	77/100	77%	
pDB2786	*Xcm*I	C-terminal FLP Mutant (FLP 1-384, plus 14 other residues)	B	0.079	73/100	73%	72%
pDB2786	*Xcm*I	C-terminal FLP Mutant (FLP 1-384, plus 14 other residues)	B	0.052	61/100	61%	
pDB2823	*Xcm*I	FLP (1-382)	B	0.071	70/100	70%	64%
pDB2823	*Xcm*I	FLP (1-382)	B	0.055	57/100	57%	

Figure 33D

Table 3 (part 4)

Plasmid	Insertion Site	Insertion Site Details	2µm Form	Initial OD_{600}	Leucine Prototrophs		
					Number	%	Average
pDB2813	*Hga*I	Inverted Repeat	A	0.057	100/100	100%	100%
pDB2813	*Hga*I	Inverted Repeat	A	0.076	100/100	100%	
pDB2808	*Fsp*I	Inverted Repeat	A	0.058	100/100	100%	100%
pDB2808	*Fsp*I	Inverted Repeat	A	0.060	100/100	100%	
pDB2812	*Hga*I	Inverted Repeat	B	0.062	100/100	100%	100%
pDB2812	*Hga*I	Inverted Repeat	B	0.071	100/100	100%	

Figure 34

SEQ ID NO:1

```
   1 ATGGACGACA TTGAAACAGC CAAGAATCTG ACGGTAAAAG CACGTACAGC TTATAGCGTC TGGGATGTAT GTCGGCTGTT TATTGAAATG ATTGCTCCTG
 101 ATGTAGATAT TGATATAGAG AGTAAACGTA AGTCTGATGA GCTACTCTTT CCAGGATATG TCATAAGGCC CATGGAATCT CTCACAACCG GTAGGCCGTA
 201 TGGTCTTGAT TCTAGCGCAG AAGATTCCAG CGTATCTTCT GACTCCAGTG CTGAGGTAAT TTTGCCTGCT GCGAAGATGG TTAAGGAAAG GTTTGATTCG
 301 ATTGGAAATG GTATGCTCTC TTCACAAGAA GCAAGTCAGG CTGCCATAGA TTTGATGCTA CAGAATAACA AGCTGTTAGA CAATAGAAAG CAACTATACA
 401 AATCTATTGC TATAATAATA GGAAGATTGC CCGAGAAAGA CAAGAAGAGA GCTACCGAAA TGCTCATGAG AAAAATGGAT TGTACACAGT TATTAGTCCC
 501 ACCAGCTCCA ACGGAAGAAG ATGTTATGAA GCTCGTAAGC GTCGTTACCC AATTGCTTAC TTTAGTTCCA CCAGATCGTC AAGCTGCTTT AATAGGTGAT
 601 TTATTCATCC CGGAATCTCT AAAGGATATA TTCAATAGTT TCAATGAACT GGCGGCAGAG AATCGTTTAC AGCAAAAAAA GAGTGAGTTG GAAGGAAGGA
 701 CTGAAGTGAA CCATGCTAAT ACAAATGAAG AAGTTCCCTC CAGGCGAACA AGAAGTAGAG ACACAAATGC AAGAGGAGCA TATAAATTAC AAAACACCAT
 801 CACTGAGGGC CCTAAAGCGG TTCCCACGAA AAAAAGGAGA GTAGCAACGA GGGTAAGGGG CAGAAAATCA CGTAATACTT CTAGGGTATG ATCCAATATC
 901 AAAGGAAATG ATAGCATTGA AGGATGAGAC TAATCCAATT GAGGAGTGGC AGCATATAGA ACAGCTAAAG GGTAGTGCTG AAGGAAGCAT ACGATACCCC
1001 GCATGGAATG GGATAATATC ACAGGAGGTA CTAGACTACC TTTCATCCTA CATAAATAGA CGCATATAAG TACGCATTTA AGCATAAACA CGCACTATGC
1101 CGTTCTTCTC ATGTATATAT ATATACAGGC AACACGCAGA TATAGGTGCG ACGTGAACAG TGAGCTGTAT GTGCGCAGCT CGCGTTGCAT TTTCGGAAAC
1201 GCTCGTTTTC GGAAACGCTT TGAAGTTCCT ATTCCGAAGT TCCTATTCTC TAGAAAGTAT AGGAACTTCA GAGCGCTTTT GAAAACCAAA AGCGCTCTGA
1301 AGACGCACTT TCAAAAAACC AAAAACGCAC CGGACTGTAA CGAGCTACTA AAATATTGCG AATACCGCTT CCACAAACAT TGCTCAAAAG TATCTCTTTG
1401 CTATATATCT CTGTGCTATA TCCCTATATA ACCTACCCAT CCACCTTTCG CTCCTTGAAC TTGCATCTAA ACTCGACCTC TACAT
```

Figure 35

SEQ ID NO:2

```
   1 ATGCCACAAT TTGGTATATT ATGTAAAACA CCACCTAAGG TGCTTGTTGG TCAGTTTGTG GAAAGGTTTG AAAGACCTTC AGGTGAGAAA ATAGCATTAT
 101 GTGCTGCTGA ACTAACCTAT TTATGTTGGA TGATTACACA TAACGGAACA GCAATCAAGA GAGCCACATT CATGAGCTAT AATACTATCA TAAGCAATTC
 201 GCTGAGTTTC GATATTGTCA ATAAATCACT CCAGTTTAAA TACAAGACGC AAAAAGCAAC AATTCTGGAA GCCTCATTAA AGAAATTGAT TCCTGCTTGG
 301 GAATTTACAA TTATTCCTTA CTATGGACAA AAACATCAAT CTGATATCAC TGATATTGTA AGTAGTTTGC AATTACAGTT CGAATCATCG GAAGAAGCAG
 401 ATAAGGGAAA TAGCCACAGT AAAAAAATGC TTAAAGCACT TCTAAGTGAG GGTGAAAGCA TCTGGGAGAT CACTGAGAAA ATACTAAATT CGTTTGAGTA
 501 TACTTCGAGA TTTACAAAAA CAAAAACTTT ATACCAATTC CTCTTCCTAG CTACTTTCAT CAATTGTGGA AGATTCAGCG ATATTAAGAA CGTTGATCCG
 601 AAATCATTTA AATTAGTCCA AAATAAGTAT CTGGGAGTAA TAATCCAGTG TTTAGTGACA GAGACAAAGA CAAGCGTTAG TAGGCACATA TACTTCTTTA
 701 GCGCAAGGGG TAGGATCGAT CCACTTGTAT ATTTGGATGA ATTTTTGAGG AATTCTGAAC CAGTCCTAAA ACGAGTAAAT AGGACCGGCA ATTCTTCAAG
 801 CAATAAACAG GAATACCAAT TATTAAAAGA TAACTTAGTC AGATCGTACA ATAAAGCTTT GAAGAAAAAT GCGCCTTATT CAATCTTTGC TATAAAAAAT
 901 GGCCCAAAAT CTCACATTGG AAGACATTTG ATGACCTCAT TTCTTTCAAT GAAGGGCCTA ACGGAGTTGA CTAATGTTGT GGGAAATTGG AGCGATAAGC
1001 GTGCTTCTGC CGTGGCCAGG ACAACGTATA CTCATCAGAT AACAGCAATA CCTGATCACT ACTTCGCACT AGTTTCTCGG TACTATGCAT ATGATCCAAT
1101 ATCAAAGGAA ATGATAGCAT TGAAGGATGA GACTAATCCA ATTGAGGAGT GGCAGCATAT AGAACAGCTA AAGGGTAGTG CTGAAGGAAG CATACGATAC
1201 CCCGCATGGA ATGGGATAAT ATCACAGGAG GTACTAGACT ACCTTTCATC CTACATAAAT AGACGCATAT AAGTACGCAT TTAAGCATAA ACACGCACTA
1301 TGCCGTTCTT CTCATGTATA TATATATACA GGCAACACGC AGATATAGGT GCGACGTGAA CAGTGAGCTG TATGTGCGCA GCTCGCGTTG CATTTTCGGA
1401 AGCGCTCGTT TTCGGAAACG CTTTGAAGTT CCTATTCCGA AGTTCCTATT CTCTAGAAAG TATAGGAACT TCAGAGCGCT TTTGAAAACC AAAAGCGCTC
1501 TGAAGACGCA CTTTCAAAAA ACCAAAAACG CACCGGACTG TAACGAGCTA CTAAAATATT GCGAATACCG CTTCCACAAA CATTGCTCAA AAGTATCTCT
1601 TTGCTATATA TCTCTGTGCT ATATCCCTAT ATAACCTACC CATCCACCTT TCGCTCCTTG AACTTGCATC TAAACTCGAC CTCTACAT
```

Figure 36